



ELSEVIER

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

The influence of visual feedback and gender dynamics on performance, perception and communication strategies in CSCW

Theodora Koulouri^{a,*}, Stanislao Lauria^b, Robert D. Macredie^b^a Department of Informatics, King's College London, WC2R 2LS London, UK^b Department of Computer Science, Brunel University, UB8 3PH Uxbridge, UK

ARTICLE INFO

Article history:

Received 25 September 2015

Received in revised form

16 August 2016

Accepted 4 September 2016

Available online 6 September 2016

Keywords:

Computer-supported cooperative work

Collaboration

Gender

Visual feedback

Shared workspace

Virtual Environment navigation

ABSTRACT

The effects of gender in human communication and human-computer interaction are well-known, yet little is understood about how it influences performance in the complex, collaborative tasks in computer-mediated settings – referred to as Computer-Supported Collaborative Work (CSCW) – that are increasingly fundamental to the way in which people work. In such tasks, visual feedback about objects and events is particularly valuable because it facilitates joint reference and attention, and enables the monitoring of people's actions and task progress. As such, software to support CSCW frequently provides shared visual workspace. While numerous studies describe and explain the impact of visual feedback in CSCW, research has not considered whether there are differences in how females and males use it, are aided by it, or are affected by its absence. To address these knowledge gaps, this study explores the effect of gender – and its interactions within pairs – in CSCW, with and without visual feedback. An experimental study is reported in which mixed-gender and same-gender pairs communicate to complete a collaborative navigation task, with one of the participants being under the impression that s/he is interacting with a robot (to avoid gender-related social preconceptions). The study analyses performance, perceptions and communication strategies. As predicted, there was a significant benefit associated with visual feedback in terms of language economy and efficiency. However, it was also found that visual feedback may be disruptive to task performance, because it relaxes the users' precision criteria and inflates their assumptions of shared perspective. While no actual performance difference was found between males and females in the navigation task, females rated their own performance less positively than did males. In terms of communication strategies, males had a strong tendency to introduce novel vocabulary when communication problems occurred, while females exhibited more conservative behaviour. When visual feedback was removed, females adapted their strategies drastically and effectively, increasing the quality and specificity of the verbal interaction, repeating and re-using vocabulary, while the behaviour of males remained consistent. These results are used to produce design recommendations for CSCW systems that will suit users of both genders and enable effective collaboration.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

While Computer-Supported Cooperative Work (CSCW) is steadily becoming the norm for many collaborative activities, there is limited understanding about the factors that influence its success. Research has suggested the influence of numerous factors pertaining to, inter alia, the individual, the group, the situation, the task and the features/affordances of the mediating technology (Patel et al., 2012). At the same time, the interplay between factors is likely to generate interaction effects while varying the individual effects. The study reported in this paper focuses on two factors and their interactions: the use of visual feedback as a resource for CSCW; and the gender of the individuals involved.

* Corresponding author.

E-mail address: theodora.koulouri@kcl.ac.uk (T. Koulouri).

not for brainstorming tasks (Whittaker et al., 1993), and is more useful for linguistically complex tasks than for simple ones (Ranjan et al., 2007; Gergle et al., 2004). Second, it depends on the features of the technology; that is, the kind and amount of visual feedback provided. For example, sharing workspace (being able to view physical actions and movements and relevant shared objects in the environment) is more important than being able to observe each other's faces or bodies (Whittaker, 2003a; Fussell et al., 2003a, 2003b; Kraut et al., 2003; Anderson et al., 2000). Moreover, Sellen (1995) found that visual feedback improves satisfaction, but not performance. In fact, visual feedback can also act disruptively as, for example, in cases of even slight delays in, or lack of, speech/visual feedback synchronicity (Gergle et al., 2006, 2004; O'Malley et al., 1996). A shared workspace has been argued to create inflated assumptions of common perspective, ultimately leading to coordination problems (Whittaker, 2003a; Schober, 1993).

More recently, research has also focused on how group and individual user characteristics moderate the benefit of visual feedback (see, Fussell and Setlock (2014), for a review); for example, studies have explored the effects of the expertise and cultural background of users. In particular, visual aids to communicate information may be more important in expert-novice interactions than in interactions between users with the same level of expertise, given that novices often lack the ability to use and understand domain-related terminology (Bromme et al., 2005). More specifically, availability of gaze and gesture cues are found to be more valuable to novices than to experts (Anderson et al., 2007; Stein and Brennan, 2004). Moreover, visual feedback has been found to provide no benefit to users of non-Western cultures (Setlock et al., 2004, 2007). As such, these findings motivate further research regarding the mechanisms, conditions and factors (relating to the technology as well as the user) that make visual feedback a valuable resource for CSCW.

The second factor considered in the study is gender. Studies in diverse fields, ranging from psychology and linguistics to business and computing, have suggested that males and females communicate and process information differently (for instance, see Halpern (2000), Beckwith and Burnett (2004)). In particular, gender has been found to be a major factor underlying cognitive abilities, such that males outperform females in the majority of visuo-spatial tasks, while females perform better in most verbal ability tests (Halpern et al., 2007; Ullman et al., 2008). In addition to differences in performance, qualitative differences are observed in terms of communication style, use of linguistic elements and level of participation in both settings of face-to-face and computer-mediated communication (see, for example, Fischer (2011), Herring (2000), Herring and Stoerger (2014)). However, a review of meta-analyses of gender literature revealed that contextual factors influence the magnitude of gender differences, presenting the theoretical argument that gender differences may be moderated, exacerbated or even reversed owing to dyadic interactions between participants (Hyde, 2005).

The notion that interactive communication itself shapes performance and language use – either amplifying or offsetting gender differences – lies at the heart of the study reported in this paper. It originates from empirical research on dialogue (work within the Interactive Alignment Model (Pickering and Garrod, 2004) and the Collaborative Model (Clark, 1996)), which postulates that the way in which people produce and understand language, and coordinate in task-oriented dialogue, largely depends on inter-individual processes and the context of use. In particular, a well-known phenomenon is linguistic alignment – the tendency of speakers to adapt to each other's pronunciation, word and syntax choices – a mechanism/strategy which is argued to underlie communication success. If, indeed, communication success is linked to linguistic alignment, it is important to discover whether alignment is mediated by gender (i.e., whether female or male speakers have stronger tendencies to align to their partners).

Finally, literature in the field of computing has recognised that there are notable differences in the ways in which females and males interact through and with technology. These differences pertain to skills, performance outcomes, perceptions and attitudes across numerous domains of Human-Computer Interaction (HCI) (Chen and Macredie, 2010). Specifically in the area of CSCW, initial evidence suggests that gender composition of groups and pairs of adults (Richert et al., 2011; Sun, 2008) and children (Underwood et al., 2000) affects computer-based collaboration in terms of performance and communication strategies. Despite these findings, our understanding of the ways in which gender interacts with the characteristics of the technology and mediates its effectiveness and acceptance remains incomplete (Burnett et al., 2011). System designs that exclude consideration of gender and other human factors ultimately lead to systems that marginalise the needs and preferences of large user groups (Dillon and Watson, 1996). As such, focused research is needed to guide system designs that can support users of both genders in computer-mediated collaboration. Moreover, in light of the evidence that visual feedback benefits collaboration and communication, the question that naturally follows is whether its utility, too, is mediated by gender.

This study builds on insights, and addresses questions, that have emerged from previous work by the authors. Koulouri et al. (2012) found that performance in navigation tasks does not depend on the individual's gender, but on the interaction of genders (that is, an individual's ability to provide/follow route instructions is also contingent on the gender of their partners). In addition, the study provided initial support to the notion that, while people may have their gender-preferential style of giving route instructions, they are willing to adapt it to suit their addressee's needs. These findings emphasised the importance of interaction and adaptation processes in collaborative tasks. Drawing on this earlier research, the study reported in this paper looks at actual as well as perceived performance in the navigation task, and undertakes fine-grained linguistic analysis to characterise and 'quantify' communication strategies and linguistic alignment as a medium for communication success. In addition, it extends the experiment to include two different interaction conditions in order to describe and explain how female and male partners benefit from visual feedback, or are affected by its absence, in CSCW.

The remainder of the paper is organised as follows. First, the area in which the research is situated is defined: Section 2 builds on past research and frames 10 research hypotheses targeting the effects of gender, visual feedback, and their interaction effects on computer-mediated collaboration. Section 3 presents the experimental methodology developed to address the research hypotheses, which involved pairs of participants collaborating in a simulated robot navigation task, and details the metrics used to capture performance and perceptions, as well as to describe user communication strategies. Section 4 reports the results of the statistical analysis on the metrics, which revealed simple and interaction effects. Section 5 discusses the findings of the study in light of existing literature and ends with recommendations for CSCW design. Section 6 reviews the limitations of the research and outlines additional areas that merit further exploration. Section 7 provides brief concluding remarks. The terms *visual information*, *shared workspace* and *visual feedback* will be used interchangeably in the rest of the paper to refer to shared visual workspace.

2. Background and research hypotheses

The previous section introduced the strands of research, from which the two main theses of the study are derived: (i) that there are differences in how females and males use and experience technology, which also emerge in CSCW; (ii) that visual feedback in CSCW may lead to more successful collaborative work. Drawing on existing literature and its findings, this section generates 10 research

hypotheses, which are grouped into three subsections corresponding to the gender factor, the visual feedback factor, and their interaction effects. Given the study's HCI scope, gender and visual feedback effects are evaluated using the concepts of effectiveness (defined in terms of error rates), efficiency (that is, resources expended, such as time and word count), and user perceptions of the interaction. For brevity, the term performance is used to mean both efficiency and effectiveness. Finally, the study assesses communication behaviour and strategies through discourse analysis.

2.1. The effect of gender

Gender, and its effect in relation to cognitive abilities and communication styles, has a broad impact on computer skills and technology requirements. In particular, research has revealed gender differences in usage, preferences and perceptions in various application areas of HCI, including online shopping and web applications (Bae and Lee, 2011; Bimber, 2000), computer games (Cassell, 1998; Hartmann and Klimmt, 2006), virtual environments (Cutmore et al., 2000; Yoon et al., 2015), car navigation systems (Lin and Chien, 2010), office software suites (Burnett et al., 2011), domestic appliances (Blackwell et al., 2009) and decision support systems (Djamasbi and Loiacono, 2008). Existing literature has also discussed the role of gender in the areas of social computer-mediated communication (CMC) (on a larger scale, see, for example, Fischer (2011), Herring (2000), Herring and Stoerger (2014)) and CSCW (presented in this section), providing some evidence that gender pair/group composition impacts performance outcomes and experience. Given the growing popularity of CMC and CSCW systems, further systematic work is needed to clarify how gender mediates the use of such systems and influences group dynamics and collaborative work. This section discusses previous research, and the ensuing hypotheses concern the effect of gender (and gender composition) on performance, user perceptions and communication strategies in CSCW. Owing to the lack of conclusive evidence, non-directional hypotheses are constructed.

Studies have addressed whether gender and pair/group composition influence performance. Community co-membership (that is, belonging to the same group in terms of gender, age, culture, geographical region, etc.) reinforces the common ground (Clark and Marshall, 1981; Setlock et al., 2004), and could lead to a performance advantage for same-gender pairs/teams. A study in which physics students used a computer-supported collaborative learning environment found that females in all-female pairs outperformed females in mixed pairs, while males did equally well with female or male partners (Ding et al., 2011). In a day-trader collaborative game, all-female pairs were faster than all-male pairs (Sun, 2008), but the study did not include mixed-gender pairs. Prinsen et al. (2009) argued that females performed better in a collaborative learning task, because they could rely more on verbal skills and asked more questions. The authors concluded that computer-supported collaborative learning environments may have a greater utility for females because they are able to show their potential and use their linguistic skills more easily than in face-to-face interactions, in which females may face anxiety and the stereotype threat (Shih et al., 2002; Schmader et al., 2008). Stereotype threat is the popular belief about the competencies/deficiencies of a group (ethnic, racial, gender). It leads to suboptimal performance among the individuals in the group, caused by excessive cognitive and memory demands, anxiety and fear (Schmader et al., 2008). Motivated by these findings, the first research hypothesis is formulated as follows:

H_{G1}: Gender has an effect on performance.

Gender has been found to influence user perceptions, satisfaction and self-efficacy (Busch, 1995; Bao et al., 2013; Durnell

et al., 2000; Ong and Lai, 2006; Kling et al., 1999). In the domain of CSCW, Bernard et al. (2000) showed that females reported lower levels of self-efficacy and satisfaction and higher anxiety in mixed-gender groups. Savicki et al. (1996) found that all-female groups expressed greater satisfaction and confidence with group processes and communication than mixed and all-male groups. Therefore, the second hypothesis investigates whether there are perception differences between females and males in terms of their performance when collaborating to complete a task:

H_{G2}: Gender has an effect on user perceptions of performance.

Gender differences in terms of communication styles have been found in socially-oriented CMC (Savicki and Kelley, 2000; Choi et al., 2009), as well as in CSCW and computer-supported collaborative learning (CSCL) environments (Prinsen et al. 2007; Ding et al., 2011). Studies have suggested that males dominate CMC discussions, sending more messages (Carr et al., 2004), and females appear to adopt a more cooperative and agreeable conversational style (Sun, 2008). Females also ask more questions than males in CSCL situations (Ding et al., 2011; Prinsen et al., 2009). In a collaborative architectural task using a table-top interface, females tended to explicitly question and state requirements, resources and plans for action, while males tended to execute more and negotiate the plan less (Richert et al., 2011). As such, the following two hypotheses focus on the gender effect on communication behaviour, making use of measures of the frequency of queries, acknowledgements, and clarifications, and the level of specificity of the utterances, provided by both partners:

H_{G3}: Gender has an effect on communication structure, in terms of frequency of queries, acknowledgements, and clarifications.

H_{G4}: Gender has an effect on communication content, in terms of language specificity.

Linguistic alignment occurs in dialogue at all levels – phonetic, phonologic, lexical, syntactic, semantic and pragmatic – and it is argued to be the mechanism that makes communication 'easy', efficient and effective (Garrod and Pickering, 2004; Pickering and Garrod, 2004). In respect of lexical alignment, dialogue is full of repetition of the same words (Brennan and Clark, 1996; Tannen, 1989); interlocutors align in terms of vocabulary in the sense that they use the same referring expressions (Garrod and Anderson, 1987). The argument that alignment underlies communication success has been explained in either social or utilitarian terms; that is, people adapt to each other linguistically in order to be positively perceived (Giles et al., 1991), or because it increases the probability of being understood (Brennan and Clark, 1996; Pickering and Garrod, 2004). The latter interpretation was constructed through experiments with collaborative dialogues, which corresponds to the domain of this study. Females appear to have stronger tendencies to adapt their language in terms of phonetics and stylistic elements in social conversations (Fitzpatrick et al., 1995; Stupka, 2011; Namy et al., 2002; Coates, 1986; Bilous and Krauss, 1988; Mulac et al., 1988; Holmes, 1990). However, there is no empirical data to suggest that females align more than males (or vice versa) in collaborative interactions or at other linguistic levels. The following hypothesis aims to address this knowledge gap and focuses on alignment in terms of vocabulary:

H_{G5}: Gender has an effect on communication behaviour, in terms of linguistic adaptation (lexical alignment).

2.2. The effect of visual feedback

The second set of research hypotheses targets the simple effects of the absence/presence of visual feedback on performance and communications strategies in CSCW. Previous studies (Gergle et al., 2004, 2013; Kraut et al., 2002a, 2003; Clark and Krych, 2004;

Brennan, 2005) have provided a body of robust evidence about the effect of visual feedback, facilitating the formulation of specific directional hypotheses. The studies explain these effects in terms of *situation awareness* and *grounding*, largely relying on concepts from the Collaborative Model developed by Clark and his colleagues (Clark, 1996).

Research in CSCW has emphasised the role of situation awareness for the success of computer-mediated collaborations (Carroll et al., 2003; Belkadi et al., 2013). Visual feedback is fundamental in supporting situation awareness (Dourish and Bellotti, 1992; Carroll et al., 2006; Daly-Jones et al., 1998; Kraut et al., 2003; Wu et al., 2013) by enabling collaborators to monitor task status and their partner's activities; speakers can assess the progress of the task, and the information necessary towards its accomplishment. Moreover, monitoring actions and their completion means that the next instruction will be provided precisely at the required moment. Similarly, an incorrect execution is readily recognised by the partner and he/she can take immediate action towards repairing it. Drawing on these findings, the following hypothesis is formulated:

H_{v1}: *Visual feedback benefits performance.*

When interlocutors introduce and accept information, they perform a coordination process known as *grounding* (Clark and Marshall, 1981), in which they mutually establish that what has been said has also been understood. Visual feedback of understanding is faster and more secure than spoken claims of understanding (Clark and Marshall, 1981), and, as a result, grounding is not performed with discrete utterances, but with visible physical actions. Simply put, if someone following instructions is aware that their partner can see what they are doing, their action serves to demonstrate understanding and substitutes verbal turns. In this case, the speaker assumes responsibility to assess the perceptual evidence provided by the addressee. Generally, the responsibility falls to whoever is judged to have the strongest evidence, so that collective effort is minimised (Clark and Wilkes-Gibbs, 1986). However, when visual feedback is unavailable, both speakers are tasked to verbally assert that something was understood or executed, leading to different turn-taking patterns. Moreover, when sharing visual information and viewing the same objects, those objects are part of the common ground, and joint attention and reference can be easily established. As a result, the act of referring to elements in the environment is less effortful, leading to highly elliptical and deictic expressions, such as 'turn here' or 'put that there'.

To summarise this research, visible physical actions provide immediate and robust evidence of execution and understanding, which renders verbal turns redundant. Moreover, the availability of visual feedback facilitates joint reference, which enables the use of shorter referring expressions. These phenomena lead to the prediction that, when visual information is shared, communication will be more economical compared to a verbal-only condition. This is investigated through the following two specific hypotheses:

H_{v2}: *Visual feedback enables pairs to complete the task using lower frequency of queries, acknowledgements, and clarifications.*

H_{v3}: *Visual feedback enables pairs to complete the task with lower language specificity.*

2.3. The interaction effects of gender and visual feedback

The third set of research hypotheses focuses on the interaction effects of visual feedback and gender on performance and communication strategies – whether visual feedback moderates (changes the strength or the direction of) the effect of gender on performance and communication. Specifically, the first research hypothesis addresses whether the performance of females or males is more adversely

affected in response to a visually-impooverished interaction condition:

H_{GV1}: *Visual feedback moderates the performance of males and females.*

As discussed above, in addition to an impact on performance, the presence/absence of visual feedback is expected to influence the content and structure of communication. However, there is no empirical data with regards to whether females or males adapt their communication strategies more drastically, in response to a less optimal condition, without visual cues. As such, the following non-directional hypothesis is formulated:

H_{GV2}: *Visual feedback moderates the communication strategies of males and females.*

3. Methodology

The experimental study involved same-gender and mixed-gender pairs collaborating to complete a navigation task. Each pair consisted of an instructor and a follower, with the former being under the impression that he/she was instructing and interacting with a robot. The robot simulation aimed to increase the barriers to express social and personal identities, and, as such, it enabled the isolation of gender effects that arise naturally and not as a result of social bias. The literature and assumptions that motivated this setup are presented below. The associated limitations and trade-offs are discussed in Section 3.1.4.

Individuals use 'shortcut heuristics' – salient social categories, like race or gender – to form impressions about their interlocutors, especially when the situation precludes other cues and background knowledge. Such judgments about the speaker are argued to be more influential in communication than the actual information content of the message (Koh and Sundar, 2010; Zimbardo and Leippe, 1991; Norman, 1976). Moreover, gender roles and prescriptions have been found to exert a strong influence in how people behave and interact (Mazei et al., 2015; Gladstone and O'Connor, 2014). A study by Matheson (1991) investigated the extent of social perceptions relating to gender in CMC. In this study, males and females participated in a negotiation task, with the gender of their interlocutor either being known or unknown to them. In reality, however, the interlocutor was a software programme that provided standardised responses. When participants believed that their interlocutor had a gender, the 'female' interlocutor was perceived to be more cooperative. Matheson (1991) concludes that CMC interactions are largely mediated by social cues, and are influenced by stereotypes and *a priori* expectations. In addition to peer stereotyping, Herring (2000) suggested that males and females self-stereotype, and adopt the communication behaviour expected of their gender in CMC, even when gender is concealed. Finally, the performance of females has been found to deteriorate when males are present (Picucci et al., 2011; Inzlicht and Ben-Zeev, 2000). As discussed in Section 2.1, the effects of 'stereotype threat' and anxiety in performance are exacerbated when individuals carry out computing or spatial tasks (Shih et al., 2002; Schmader et al., 2008; Lawton, 1996, 1994; Beckwith, 2007).

These findings indicate that social biases associated with gender have a strong and confounding effect on language and performance. As such, by masking the gender through a robot simulation, it was possible to observe naturally-occurring cognitive processes and behaviours with less interference from social aspects of gender.

3.1. Experimental method

3.1.1. Task

The domain used in the experiment was pedestrian navigation in a simulated town, with the instructor having to guide a 'robot' (hereafter referred to as follower) to six designated locations. The

environment consisted of highly salient landmarks such as buildings, and landmarks of lower salience such as pathways, which aimed to approximate a realistic urban environment. Environments containing a fair number of landmarks have been shown to be appropriate for users of both genders (Sandstrom et al., 1998; Saucier et al., 2002; Cutmore et al., 2000). Moreover, the environment was two-dimensional, given that navigation in 3D virtual environments has been found to be cognitively demanding for most people (Cockburn and McKenzie, 2002). Navigation tasks, including 'maze game' (see, for example, Garrod and Anderson (1987) and Mills (2007)) and 'map task' scenarios, (for example, Sanford et al., 2003; O'Malley et al., 1996; Anderson et al., 1991) are part of well-known experimental paradigms in (computer-mediated) human communication research as they enable observation of natural co-operative behaviour, while the information available to participants remains finite and controlled at any point in the dialogue. The co-operative nature of the task lay in two aspects. First, in each pairing only the instructor knew the destinations and had a global view of the environment, so the follower had to rely on the instructor's route descriptions. Second, the instructor also needed the follower's descriptions to determine the robot's exact position and perspective. Data were captured on each participant's actions and utterances, to support analysis and understanding of how the participants approached the task and any problems that arose. The tasks in this study did not require learning the route through navigation or recalling the map or instructions from memory, which would give rise to different cognitive demands and errors.

3.1.2. The system

The experiment relied on a custom-built system which supported the interactive simulation and enabled real-time direct text communication between the instructor and follower. The system kept a log of the dialogues and also recorded the coordinates of the current position of the robot at the moment messages were transmitted. Thus, it was possible to analyse the descriptions against a matching record of the robot's position and reproduce its path with temporal and spatial accuracy. The interfaces used by the participants consisted of a graphical display and an instant messaging facility (the dialogue box). The dialogue box displayed each participant's messages (in green) in the upper part of the dialogue box; the messages sent by the other participant in the pair were displayed (in magenta) in the lower part of the dialogue box. The desktop PCs used by the participants were equipped with 17-in. LCD monitors with 1024 × 768 pixel resolution. The underlying design principle of the system was that the interfaces remained basic and feature-light, and were operated through simple controls, so that any performance and

communication differences would not be caused by the properties of the interfaces. The interface seen by the instructor displayed the full map of the simulated town. In order to explore the effect of the provision of visual feedback, there were two variants of the instructor's screen. The design and rationale of the visual feedback functionality, and the follower's interface, are presented below.

In the 'Visual Feedback condition', a small 'display' was available in the upper right corner of the screen showing the robot's immediate locality, but not the robot itself. This meant that the instructor shared the same visual space as the follower. This experimental decision follows the relevant literature investigating the effects of visual information (for example, Whittaker, 2003b; Kraut et al., 2003), in which the instructor can see what the follower is seeing and doing, but not the follower himself/herself. This is traced back to the 'What You See Is What I See' paradigm in the design of groupware and CSCW systems, developed by Xerox PARC (Stefik et al., 1987). The size of the 'display window' on the instructor's computer screen was approximately 7.2 × 7.2 cm; this was considered appropriate given that it displayed a scaled-down, high-fidelity image of a relatively uncluttered environment, which was also part of the instructor's own map. No delays were noted in the display of the messages or the visual feedback. Displaying the map and the follower's visual space on one screen was considered more usable and less distracting for instructors than requiring them to view two different media (for instance, paper and computer monitor or two separate monitors). This, however, resulted in a compromise in the size of the display window showing the follower's visual space. Similar interfaces have been used in the related studies by Kraut, Gergle and Fussell, discussed in Section 2.2. For example, in Kraut et al. (2003), the instructor's display consisted of the repair manual and schematics of the bicycle, and a small rectangular window in the right bottom corner showing the view from the head-mounted camera of the follower.

Clark and Brennan argue that the medium shapes the structure and content of communication by imposing different costs on the grounding process (Clark and Brennan, 1991). The medium can offer (or constrain) the following communication resources: physical co-presence, visibility, co-temporality, audibility, simultaneity, sequentiality, reviewability, revisability, mobility and tangibility. CMC presents potential barriers to communication because it variably restricts these affordances. Adapting the framework to CMC, Brennan and Lockridge (2006) removed the mobility and tangibility affordances. Similarly, Kraut et al. (2002b) have proposed a refined version of the framework, which 'decomposes' the dimensions of physical co-presence. The framework that conceptualises the key affordances of a communication medium (and their definitions) is shown in Table 1. The characteristics of the interface used in this study are explained against

Table 1
A characterisation of the interface used in the study against a framework of affordances of communication media (adapted from Clark and Brennan (1991) and Kraut et al. (2002b), as presented in Fussell and Setlock (2014)).

Affordances of Media	Interface
(1)Physical co-presence dimensions:	
– Field of view: Participants can see what entities each person is oriented towards.	Only Instructor
– Spatial perspective: Participants can see their partner's perspectives toward the environment.	No
– Display symmetry: Both parties can see equivalent aspects of one another's environments.	No
– Dimensionality: Participants have a 3D view of the environment.	No
– Spatial resolution: Participants have high-quality views of the environment.	Yes
– Temporal delay: Images as received relative when they were sent.	Yes
(2)Visibility: One participant sees each other, but not what they are doing.	No
(3)Audibility: One participant can hear another.	No
(4)Co-temporality: Messages are received without delay (close to the time that they are produced and directed at addressees), permitting fine-grained interactivity.	Yes
(5)Simultaneity: Participants can send and receive messages at the same time, allowing communication in parallel.	Yes
(6)Sequentiality: Participants take turns in an orderly fashion in a single conversation at a time; one turn's relevance to another is signalled by adjacency.	No
(7)Reviewability: Messages do not fade over time.	Yes
(8)Revisability: Messages can be revised before being sent.	Yes



Fig. 1. The interface operated by the instructor in the Visual Feedback condition, with the small window in the upper right corner displaying the robot's current location. This window was absent in the No Visual Feedback condition.



Fig. 2. The interface operated by the follower.

this framework and are also summarised in Table 1.

In the 'No Visual Feedback condition', the 'display window' feature was disabled so that the instructor had no direct visual information relating to the follower's position and actions in the environment (see Fig. 1).

The follower's interface displayed a fraction of the overall environment map, showing only the surroundings of the robot's current position (see Fig. 2). The robot (signified by a red circle with a yellow 'face') was operated by the follower using the arrow keys on the keyboard. The dialogue box also displayed a history of the instructor's previous messages to the follower.

3.1.3. Participants and procedure

A total of 64 participants (32 males and 32 females) were recruited from undergraduate and postgraduate students of various departments at a UK university. The participants were randomly allocated to the two roles (instructor or follower) and to each of the Visual Feedback conditions. Previous experience in using computers was necessary, as was familiarity with instant messaging applications. No other specific computer expertise or other skill was required in order to take part in the experiment. As shown in Table 2, pairs were formed with all possible combinations of roles and gender.

Instructors and followers were seated in separate rooms

Table 2

Pair compositions and their abbreviations, used in the remainder of the paper.

Pair Composition	Abbreviation
Female instructor – Female follower	FF
Female instructor – Male follower	FM
Male instructor – Female follower	MF
Male instructor – Male follower	MM

equipped with desktop PCs, on which the respective interfaces were displayed. The participants that were assigned to be followers were fully informed about the experimental setup. The instructors were told that they would interact directly with a robot, which for practical reasons was a computer-based, simulated version of the actual robot. The instructors were given minimal information about the robot. They were informed that the robot had advanced capacity to understand and produce spatial language and learn previous routes. This aimed to reduce the likelihood of instructors inferring during the interaction that the robot was actually a person. The pairs were given no examples of, or instructions about, how to interact with each other. The pairs attempted six tasks presented in the same order; the instructor navigated the follower from the starting point (bottom right of the map) to six designated locations (pub, lab, factory, tube, Tesco, shop). The instructors were free to plan and modify the route as they wished. The destinations were selected to require either incrementally more instructions or the use of previously described routes.

3.1.4. Limitations of the methodology

The methodology of this study involved two people collaborating in a task, one of whom believed that they were interacting with a simulated robot. This section discusses the limitations that arise from this experimental approach.

As explained in Section 3, this setup allowed for the observation of inter-gender interactions while inhibiting social elements that arise in human-human communication and have an adverse, or confounding, effect on task performance and communication. However, in realistic CSCW settings the gender of participants is typically known. As such, masking the gender introduces a degree of artificiality in the study and constitutes a limitation which needs to be addressed in future work. The second limitation of the approach lies in the fact that followers were aware that their partners were people, resulting in an asymmetry of knowledge within the dyad. This is a methodological compromise as it was assumed that if the followers had not been informed of the setup, they would have given away enough cues for the instructors to suspect that they were interacting with another person. In effect, the followers acted as confederates in the study; while a confederate in similar studies is a single, trained individual, followers were also naïve participants, and they were given no guidance on what to do or say in order to ensure that the interaction was natural and spontaneous. A related problem has to do with the 'audience design' phenomenon: people produce language based on *a priori* assumptions about what their addressees might know (Isaacs and Clark, 1987; Sacks et al., 1974; Schober, 1993, 2009), and these assumptions are often influenced by 'community co-membership' (Clark and Marshall, 1981). Therefore, the question that naturally follows is whether coordination and communication were disrupted by the inability of instructors to form such assumptions. However, even if the model (set of assumptions) regarding the addressee is initially incomplete, the model is constructed dynamically, being updated on a turn-by-turn basis throughout the interaction (Clark, 1996). As such, it is argued that the simulated robot setup may not have had a significant or lasting effect on coordination and grounding processes. For example, research by Levin and colleagues (Levin et al., 2013, 2008)

indicated that people are willing to attribute human-like cognitive characteristics to robots, when users are given sufficient time to observe intentional behaviour by the robot.

3.2. Data analysis approach

The two strands of data analysis – addressing performance and communication – and their relation to specific hypotheses are explained in this section.

3.2.1. Performance analysis

The analysis of performance targets research hypotheses H_{G1} , H_{V1} , and H_{GV1} . Following HCI practice, it uses measures of effectiveness and efficiency, as also adopted in related research in human communication (e.g., Clark and Krych, 2004) and CSCW (e.g., Gergle et al., 2004; Kraut et al., 2003). The efficiency indicators used were time, and number of instructor/follower turns and words. Effectiveness was assessed through miscommunication rates. The study adapted McRoy's HCI-oriented framework and classified miscommunication into three types: execution errors; non-understandings; and incorrect instructions (McRoy, 1998). Execution errors – instances where the follower deviated from the described route – were recorded by comparing the route executed by the follower (as determined by system logs) with the route described by the instructor. Non-understandings were responses to an instruction that requested clarification or expressed an inability to interpret the instruction. Incorrect instructions were errors by the instructor, for example confusing 'left' for 'right' when providing a route instruction.¹

To examine research hypothesis H_{G2} , user perceptions were collected using a short questionnaire. After the completion of each of the six tasks, the instructors were asked to complete the questionnaire, rating their agreement with five declarative statements of opinion. The questionnaire used a Likert scale with seven levels of agreement: *strongly disagree*; *disagree*; *slightly disagree*; *neutral*; *slightly agree*; *agree*; and *strongly agree*. The items probed five different dimensions of the instructor's experience of their interaction with the follower: perceived task completion (item 1); execution accuracy (item 2); ease of use (item 3); helpfulness (item 4); and overall satisfaction (item 5). The responses were mapped to integer values between one and seven (with seven representing the highest level of agreement). The scores associated with each statement were summed for all six tasks, which resulted in a cumulative score for each statement ranging from 6 to 42. Followers were not asked to complete the questionnaire to ensure that they would be ready to respond as soon as the instructors initiated the next task.

3.2.2. Communication analysis

Evaluating and quantifying how people use language to collaborate is a non-trivial task. The approach followed existing annotation and analysis schemes to categorise the utterances of the participants and the components of these utterances, identify the degrees of alignment, and measure their associated frequencies to be statistically tested. In particular, in order to determine communication strategies and collaborative behaviour and address hypotheses H_{G3} , H_{G4} , H_{G5} , H_{V2} , H_{V3} , and H_{GV2} , two types of fine-grained analysis were performed: dialogue act analysis (following the HCRC Dialogue Structure Coding Manual by Carletta et al. (1996)); and (two forms of) component analysis of the utterances (following the definitions and frameworks developed by Denis (1997), Tenbrink and Hui (2007) and Vanetti and Allen (1988)). These analysis schemes are outlined in the following subsections.

3.2.2.1. Dialogue acts analysis: queries, acknowledgements and clarifications. The utterances were annotated using a simplified version of the HCRC Map Task move coding scheme (HCRC Dialogue Structure Coding Manual by Carletta et al. (1996)). The scheme divides the participants' dialogue moves into *initiations* (expecting a response) or *responses* that are then classified as *Instruct*, *Query*, *Clarify*, *Acknowledge*, *Reply*, *Explain*, *Check*, etc. Following the scheme's definition, the *Instruct* dialogue act refers to a command/instruction issued to the follower. The *Query*, *Clarify* and *Acknowledge* dialogue acts are directly related to hypotheses H_{G3} and H_{V2} , and are explained below through examples from the corpus.

The *Query* dialogue act covers all questions addressed to the partner which are not clarification requests, as in the instructor's turn in example 1, below. 'I' and 'F' stand for instructor and follower, respectively.

Example 1 [NMC3_T56-57]

I: Where are you?

F: I am standing facing the Post Office, with the car park on my left.

The *Acknowledge* dialogue act is a minimal sign of positive feedback. It demonstrates that a previous utterance or action was received, understood or accepted. Acknowledgements can be formulated simply, as 'OK', 'Yes' or as the follower's response in example 2, below.

Example 2 [NMC7_T35-36]

I: go to lab and walk ahead, when you see two roads take left and then keep walking for a while and take second left

F: second left taken.

The *Clarify* dialogue act is a reply to a question that contributes information over and above what was strictly asked. The difference between 'Reply' and 'Clarify' is that 'Reply' contains only the information requested. This difference is illustrated in example 3, below; the first response by the follower is classified as a 'Reply' and the second is a 'Clarify' dialogue act. The follower's response in example 1 above was also a 'Clarify'.

Example 3 [NMF2_F62-64]

I: Is the gym on your left or right?

F: The gym is on the right.

F: Brunel is on the left.

As can be seen from the definitions and examples, the dialogue acts *Acknowledge* and *Clarify* have high intrinsic value for task-oriented interactions, by grounding or supplying extra information, and were used in the analysis.

3.2.2.2. Component analysis: language specificity. The component analysis relied on established frameworks within the domain of spatial language that define the best practices for producing route descriptions and classify 'good' versus 'poor' descriptions (Vanetti and Allen, 1988; Allen, 2000; Denis, 1997; Lovelace et al., 1999; Tenbrink and Hui, 2007). An important concept within these schemes is specificity (also termed granularity), which refers to the level of specification used by a person to describe a particular situation, event or object (Tenbrink et al., 2010; Klippel et al., 2009). For example, descriptions with low specificity are simple in form, giving turn-by-turn directions to the destination, and only

¹ These miscommunication types correspond to McRoy's misunderstandings, non-understandings, and misinterpretations.

contain spatial directions. Instructions of high specificity were defined as including landmark references, which could also be anchored spatially. In particular, the schemes agree that ‘good’ descriptions contain: (i) references to 3D landmarks (locations like buildings or bridges) that provide cues for (re-)orientation; and (ii) delimiters that provide distinguishing information about actions and environmental features. Delimiters can be distance designations that specify the boundary of action (‘move forward *until* you see a car park’, ‘*from* the bridge continue straight *to* the university’, etc.) and relational terms that specify relationships between environmental features or frame of reference (‘the lab will be *on your right*’, ‘go to shop *next to* the café’, etc.). In brief, the analysis considers the relative frequencies of these two types of component that are known to contribute to the information value of utterances.

3.2.2.3. Component analysis: linguistic alignment. The study concentrated on the phenomenon of lexical alignment, which essentially manifests as the re-use of vocabulary between partners. First, alignment was measured ‘locally’; that is, by looking at the adjacency pairs in the dialogue and comparing the two utterances (termed ‘input/output matching’ in the Interactive Alignment Model by Pickering and Garrod (2004)). An adjacency pair is a sequence of two *related* utterances by two *different* speakers, such that the second utterance is a response to the first (Levinson, 1983, p. 303). An utterance was a ‘match’ (and given a score of 1 for each repeated content element), if it contained the same content word as the utterance to which it was a response. If no content word matched, a score of 0 was given and the response was noted as a ‘mismatch’. The sum of matching content words was calculated. The annotation of alignment at the adjacency pair level is exemplified through a dialogue excerpt, shown in Table 3. In this example, the instructor first produced an instruction which does not match the previous utterance by his/her partner (a mismatch and so a 0 score was given). This was immediately reformulated to repeat the exact expression used by the follower, ‘at y-shaped junction’, which is marked as containing ‘2’ matches.

Alignment is not only a local input/output matching mechanism; it also develops over the course of a dialogue, such that interlocutors rely on a working vocabulary for a dialogue’s duration. This has been described in terms of dialogue ‘routines’ or ‘pacts’ by the Interactive Alignment Model (Pickering and Garrod, 2004) and the Collaborative Model (Clark, 1996). Therefore, in addition to capturing alignment as matches at the adjacency pair level, it was necessary to measure alignment as lexical innovation. Following the approach introduced by Mills (2007), lexical innovation was determined by comparing every constituent word in an utterance to all previous words used in the dialogue. For example, the utterance ‘go into the bendy road’ leads to a backwards search in the dialogue for any previous occurrence of ‘go’, adding ‘1’ to the innovation score if not found, and ‘0’ if found, before moving on to the next word.

Table 3

Example of lexical alignment annotation at the adjacency pair level. The first instruction is a mismatch, whereas the second instruction is a match and repeats two content words.

Utterance [MC7_S142-144]	Match
F: I am at y-shaped junction.	
I: make a right.	0
I: make a right at y-shaped junction.	2

3.2.3. Experimental design and statistical analysis

A between-subjects factorial design was used that investigated the simple and interaction effects of Instructor Gender, Follower Gender, and Visual Feedback on the performance-related and

Table 4

List of measures, hypotheses and factors within the study.

	Dependent Variables	Research Hypotheses	Factors
Performance	Words/Turns Time Miscommunication (Execution errors, Non-understandings, Incorrect Instructions)	H_{G1} H_{V1} H_{GV1}	Gender Visual Feedback Gender × Visual Feedback
	User perception dimensions	H_{G2}	Gender
Communication	Queries Clarifications Acknowledgements	H_{G3}, H_{G4} H_{V2}, H_{V3} H_{GV2}	Gender Visual Feedback Gender × Visual Feedback
	Landmark references Delimiters (distance and relational terms) ‘Matches’ (lexical alignment) Novel Words (lexical alignment)	H_{G5}	Gender

communication-related dependent variables. It should be noted that the ‘task’ served as the basis of measurement (each pair completed six tasks). Table 4 summarises the dependent variables, the factors and the research hypotheses targeted. For categorical data, relationships were investigated through chi-square tests of independence. When appropriate, the chi-square analysis followed a ‘top-down’ approach in order to identify the locus of a significant association, and separate chi-square tests were performed on the Visual Feedback and No Visual Feedback data.

4. Results

4.1. Effects of gender and visual feedback on performance

To test hypotheses H_{G1}, H_{V1}, H_{GV1} , which targeted performance, the analysis considered the frequency of words and turns required to complete a task, and the frequency of miscommunication.

The analysis revealed a main effect of Visual Feedback on number of words per task ($F_{(1,24)}=6.904, p=0.015, \eta^2=0.191, d=0.94$). Pairs in the No Visual Feedback condition (Mean=99.5, SD=34.57) required a larger number of words to complete each task than the Visual Feedback pairs (Mean=72.46, SD=21.27). Both instructors and followers, individually, used a larger number of words under the No Visual Feedback condition. No significant difference in completion time was found.

The analysis of the proportion of turns by the instructor contributed interesting results. A significant main effect of Visual Feedback initially indicated that instructors in the Visual Feedback condition produced 57% of all turns, which dropped by 6% in the No Visual Feedback condition ($F_{(1,23)}=5.5, p=0.028, \eta^2=0.131, d=0.84$). This result was refined, as a significant interaction effect of Visual Feedback by Instructor Gender was found ($F_{(1,23)}=5.548, p=0.027, \eta^2=0.137$). As illustrated in Fig. 3, female instructors dominated the conversational floor in the Visual Feedback condition, having produced over 61.9% of turns. However, when visual feedback was disabled, female instructors’ turn-possession was balanced, dropping to 50.5%. Comparisons between the groups verified the difference between female instructors in the Visual Feedback condition and No Visual Feedback condition ($t_{(14)}=3.211, p=0.006, d=1.6$). In contrast, the turn ratio of male instructors remained consistent across conditions.

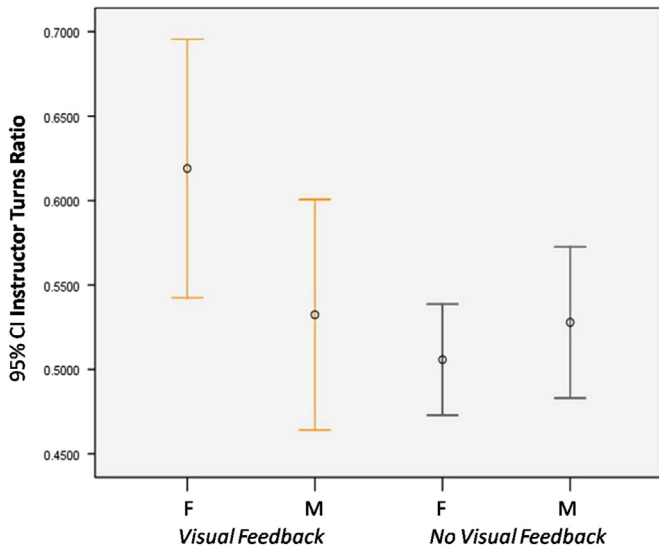


Fig. 3. The turn ratios for female and male instructors ('F' and 'M', respectively) in the Visual Feedback and No Visual Feedback conditions. The first and third error bars do not overlap, indicating significant differences between female instructors in the two conditions.

As detailed in Section 3.2.1, miscommunication caused by the instructor was estimated by the number of incorrect instructions. Follower-attributed miscommunication encompasses two measures: number of (i) execution errors; and (ii) follower turns that were tagged as expressing non-understanding.

The three-way ANOVA revealed a strong significant main effect of Visual Feedback on the number of incorrect instructions per task. Surprisingly, the number of incorrect instructions per task was close to zero in the No Visual Feedback condition and high in the condition in which the instructor could confirm at all times the actions and understanding of the follower ($F_{(1,23)}=13.784, p=0.001, \eta^2=0.304, d=1.35$). The Instructor Gender \times Follower Gender interaction was found to be significant ($F_{(1,23)}=4.797, p=0.039, \eta^2=0.106$) indicating that instructors in mixed-gender pairs (FM and MF) tended to be less accurate compared to instructors speaking to followers of the same gender (FF and MM). The contrast between same-gender and mixed-gender pairs also confirmed the finding ($t_{(29)}=-2.251, p=0.032, d=0.81$).

Similarly, the ANOVA conducted on number of non-understandings yielded a significant main effect of Visual Feedback. Interestingly, when participants shared visual information, followers produced a greater number of non-understandings ($F_{(1,24)}=4.324, p=0.048, \eta^2=0.134, d=0.76$). Finally, for execution errors as the dependent variable, no differences were found among the groups. The results are summarised in Fig. 4, which shows the distributions of incorrect instructions, non-understandings and execution errors across the two conditions.

These results address the hypotheses that relate to the effects of gender and visual feedback on performance. Fewer words were needed to complete tasks with visual feedback. However, there appears to be a trade-off with accuracy, such that miscommunication was higher when visual feedback was available. Average completion times were similar across conditions, possibly because Visual Feedback pairs expended any time advantage to resolve miscommunication. As such, the results of this analysis do not fully support hypothesis H_{V1} .

No interaction effects were observed, so hypothesis H_{CV1} cannot be confirmed. However, the difference in turn ratios suggests that females are more sensitive to changes in interaction condition, providing initial support to hypothesis H_{CV2} . Same-gender pairs also appear to be more accurate in terms of instructions, which alone may not suffice to validate hypothesis H_{C1} .

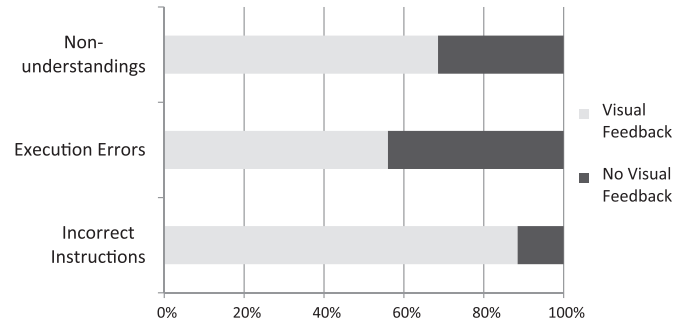


Fig. 4. Distribution of miscommunication in the Visual Feedback and No Visual Feedback conditions. The frequency of non-understandings and incorrect instructions was significantly higher when visual information was shared.

4.2. User perceptions

A mixed ANOVA design was employed to explore the effect of gender on user perceptions (hypothesis H_{G2}). The within-subjects factor was Statement, with five levels corresponding to the statements in the questionnaire. The between-subjects factors were Visual Feedback, Instructor Gender and Follower Gender. The analysis found a significant interaction effect of Instructor Gender and Statement ($F_{(3,626,79,768)}=2.750, p=0.038, \eta^2=0.084$). In particular, the results indicated that male instructors perceived higher task success than females (item 1). System accuracy was rated more favourably by female instructors (item 3). User satisfaction (item 5) was similar for both genders. System ease of use (item 2) and helpfulness (item 4) were also not significantly different. The mean summed scores for each question are shown in Fig. 5.

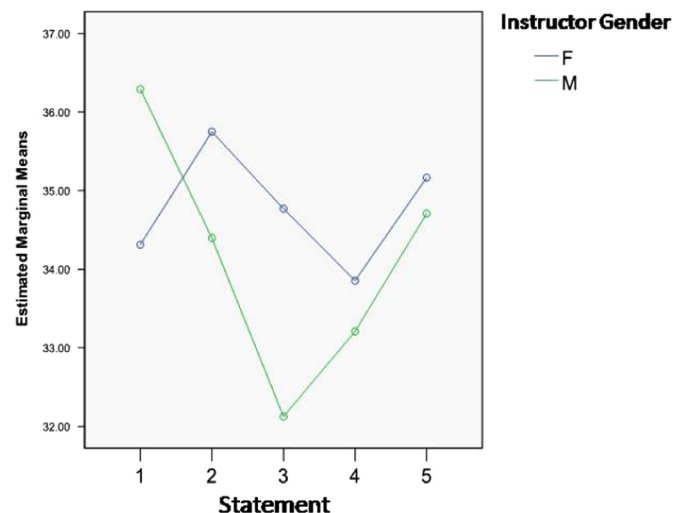


Fig. 5. Mean summed scores for each statement for female and male users. The statements were the following: 1: I did well in completing the task; 2: The system was easy to use; 3: The system was accurate; 4: The system was helpful; 5: I am generally satisfied with this interaction. Gender differences were confirmed for items 1 and 3.

This result appears to confirm hypothesis H_{G2} , which predicted that gender mediates user perceptions of task success.

4.3. Effects of gender and visual feedback on communication

In addition to their effects on task performance and perceptions, gender and visual feedback were expected to influence how instructors and followers communicated about the task. To determine communication strategies, statistical analysis was performed on the frequencies of dialogue acts and utterance

components associated with higher information value and specificity, and on the rate of vocabulary repetition and re-use.

4.3.1. Queries, acknowledgements, and clarifications

The analysis presented in this section considers the frequencies of certain dialogue acts: the number of queries (questions), acknowledgements (positive feedback to show that the utterance to which it responds has been understood and accepted) and clarifications (responses to questions that give information over and beyond what was asked) issued by the instructor and follower. The frequencies of these dialogue acts were used to address hypotheses H_{G3} , H_{V2} , and H_{GV2} .

The three-way ANOVA performed on the number of *instructor queries* yielded a significant main effect of the Visual Feedback factor ($F_{(1,22)}=14.710$, $p=0.001$, $\eta^2=0.251$, $d=1.2$). In particular, the instructor queries showed a dramatic increase in the No Visual Feedback condition. The analysis also revealed an interaction effect between Visual Feedback and Instructor Gender ($F_{(1,22)}=7.247$, $p=0.013$, $\eta^2=0.124$). *T*-tests and inspection of the error bar charts confirmed that the greatest number of queries was given by female instructors in the No Visual Feedback condition.

Finally, a significant three-way interaction of Visual Feedback by Instructor Gender by Follower Gender was detected ($F_{(1,22)}=4.203$, $p=0.05$, $\eta^2=0.072$). The presence of the three-way interaction refined the result and indicated that, although in the Visual Feedback condition female instructors paired with female followers rarely asked questions ($M=0.25$, $SD=0.29$), when visual information was not shared, the number of their queries 'exploded', increasing by 2.36 standard deviations ($M=3.45$, $SD=1.89$). The effect is illustrated in Fig. 6.

The analysis looked at the other side of the communication, the number of follower queries per task, and also revealed a main effect of Visual Feedback ($F_{(1,23)}=11.014$, $p=0.003$, $\eta^2=0.274$, $d=1.17$), but inversely: the followers issued a larger number of queries when their partners were able to monitor their actions.

The analysis of *acknowledgements* per task revealed an analogous pattern of significant effects: in the absence of shared workspace, participants produced a larger number of acknowledgements to signal understanding and acceptance of previous statements ($F_{(1,22)}=4.459$, $p=0.046$, $\eta^2=0.102$, $d=0.74$). This effect was overshadowed by a significant effect of Visual Feedback by Instructor Gender ($F_{(1,22)}=6.786$, $p=0.016$, $\eta^2=0.155$). Inspection of the error bar charts and *t*-tests showed that pairs of Female instructors in the No Visual Feedback condition provided a significantly higher number of acknowledgements compared to the other groups.

A conclusive result was reached through the second-order interaction effect of Visual Feedback by Instructor Gender by Follower Gender ($F_{(1,22)}=4.195$, $p=0.05$, $\eta^2=0.096$, $d=2.23$). In the Visual Feedback condition, FF pairs exchanged very few acknowledgements ($M=1.625$, $SD=1.5$). In contrast, in the No Visual Feedback condition, the number of acknowledgements for FF pairs quadrupled ($M=6.725$, $SD=2.86$), which translates to a difference of 2.23 standard deviations. Fig. 6 illustrates the result by showing the interaction of Instructor Gender by Follower Gender for each level of Visual Feedback.

The analysis of the number of acknowledgements by the follower also showed that when visual information is not shared, followers more frequently provide evidence of positive understanding ($F_{(1,23)}=9.629$, $p=0.005$, $\eta^2=0.22$, $d=1.04$).

The analysis of dialogue acts investigated the number of *clarifications* per task provided by the pairs. Inspection of the dialogue data showed that clarifications were provided exclusively by followers. There was a significant effect of Visual Feedback ($F_{(1,24)}=6.405$, $p=0.018$, $\eta^2=0.173$, $d=0.89$). In particular, followers gave a higher number of replies that were richer in information, in the absence of shared visual space.

Taken together, the results show that gender and visual feedback change the communication strategies used by participants to

complete the task (supporting hypotheses H_{G3} and H_{V2} , respectively). Most importantly, it was found that all-female pairs adapt drastically when the CSCW medium excludes visual feedback, providing initial support to hypothesis H_{GV2} .

4.3.2. Language specificity

The level of specificity was determined by the presence and frequency of landmarks and delimiters. This analysis served to test hypotheses H_{G4} , H_{V3} , and H_{GV2} .

The analysis on all instructor turns revealed that *landmark references* (for example, references to buildings, bridges and other salient locations) were most frequently used in the No Visual Feedback condition ($F_{(1,23)}=10.893$, $p=0.003$, $\eta^2=0.24$, $d=1.11$) and when the addressees were female ($F_{(1,23)}=4.286$, $p=0.05$, $\eta^2=0.094$, $d=0.63$). In addition, the analysis revealed an interaction effect of Visual Feedback and Instructor Gender ($F_{(1,23)}=4.598$, $p=0.043$, $\eta^2=0.101$). The interaction was explored through *t*-tests which showed differences of almost two standard deviations between female instructors under the Visual Feedback and No Visual Feedback conditions ($t_{(14)}=-3.424$, $p=0.004$, $d=1.71$). In particular, female instructors included three times as many landmark references when the follower's actions were not visible. The findings are illustrated in Fig. 7.

Mirroring their partners' behaviour, followers in the No Visual Feedback condition were also found to use more landmark references (contained in 33% of follower utterances in the Visual Feedback condition compared to 53% of utterances in the No Visual Feedback condition) ($F_{(1,24)}=8.818$, $p=0.007$, $\eta^2=0.247$, $d=0.97$). The Instructor and Follower Gender factors were not significant.

In relation to the frequency of *delimiters* in instructor utterances, the ANOVA revealed a significant main effect of Visual Feedback for boundary/distance designations. These delimiters, which specify the boundary of the route, were scarcely used in the Visual Feedback condition ($F_{(1,23)}=4.539$, $p=0.044$, $\eta^2=0.136$, $d=0.77$).

Relational terms specify the relation between speaker and an environmental feature ('on your left') or between different environmental features. Followers were found to incorporate a larger number of these terms in their utterances in the No Visual Feedback condition ($F_{(1,23)}=5.332$, $p=0.03$, $\eta^2=0.182$, $d=0.90$); that is, when not being monitored, followers tended to be explicit about the frame of reference. There was no significant effect of Instructor Gender on the frequency of relational terms.

In addition, the analysis concentrated on the frequency of simple or compound instructions. When instructions consisted of one or two components, verb or verb and direction, they were considered simple (for instance, 'walk straight ahead', 'move forward', 'turn right'). Instructions with more than two components were compound (for instance, 'walk straight ahead until you reach the road junction', which has four components). As expected, the frequency of simple instructions that contained only the verb and the direction of movement was lower in the No Visual Feedback ($F_{(1,24)}=4.769$, $p=0.039$, $\eta^2=0.144$, $d=0.77$).

A three-way interaction effect of Visual Feedback by Instructor Gender by Follower Gender was also detected ($F_{(1,24)}=4.381$, $p=0.047$, $\eta^2=0.126$). The difference was statistically significant only for FF pairs between conditions, showing that simple instructions by female instructors paired with females dramatically decreased in the No Visual Feedback condition. The interaction is plotted for each level of Visual Feedback in Fig. 8.

To sum up, these results indicate that when visual feedback is available, the pairs are able to complete the task with lower specificity, which confirms hypothesis H_{V3} . A complex gender composition effect was revealed that indicates female instructors, and all-female pairs, adapted their behaviour in order to ensure communication success in less optimal conditions, supporting hypotheses H_{G4} and H_{GV2} .

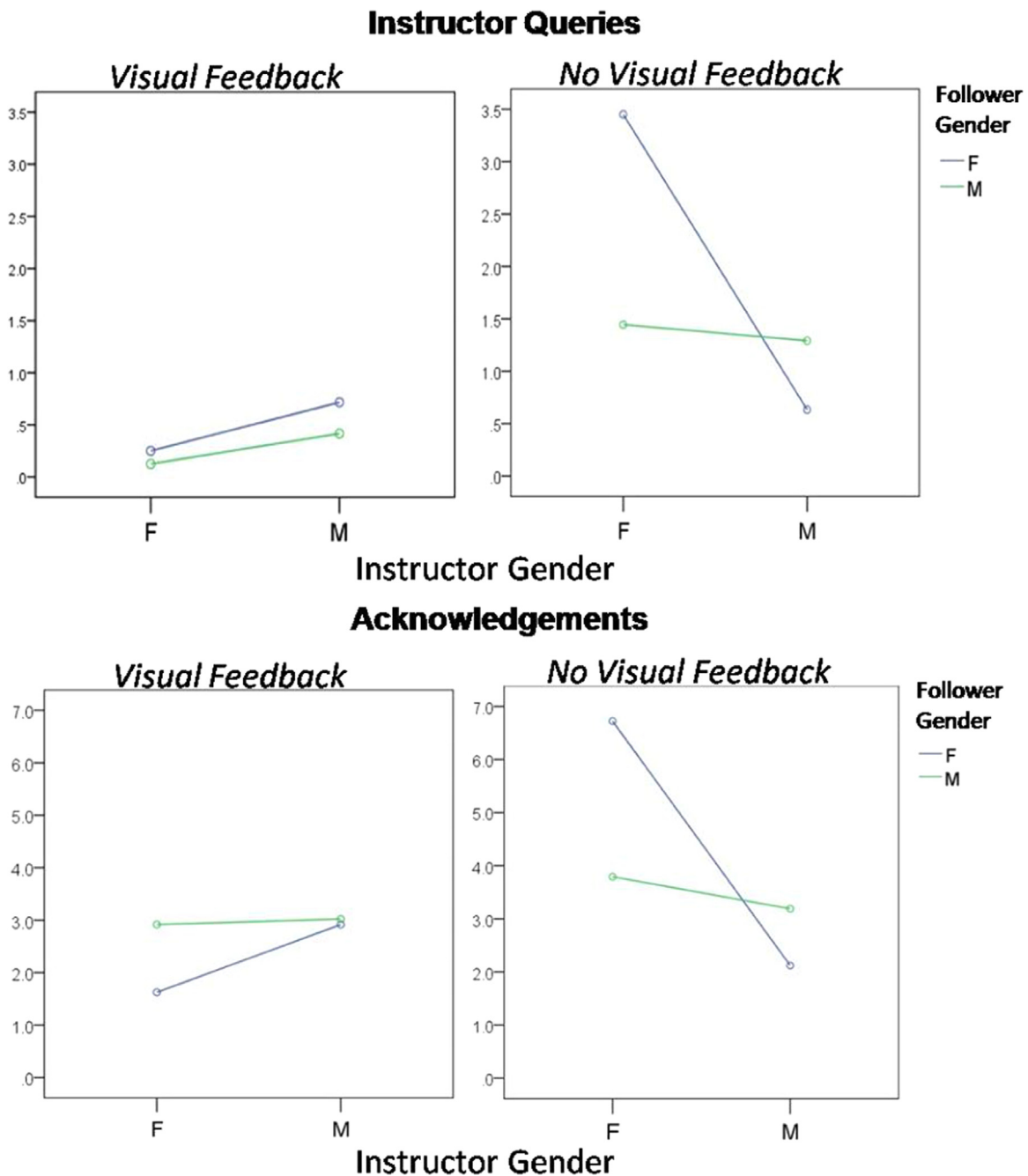


Fig. 6. Top Graphs: interaction of Instructor Gender and Follower Gender for each level of Visual Feedback. The Y axis represents the means of queries by Female or Male instructors. Bottom Graphs: interaction of Instructor Gender and Follower Gender for each level of Visual Feedback. The Y axis represents the means of acknowledgements given by Female and Male instructors. Pronounced differences are found between FF pairs in the two conditions for instructor queries and acknowledgements.

4.3.3. Linguistic alignment

To address hypothesis H_{G5} , the analysis investigated whether alignment is mediated by gender; that is, whether Instructor Gender and Follower Gender have an effect on alignment as: (i) input/output matching and (ii) lexical innovation.

First, the ANOVA failed to produce significant effects of gender on number of 'matches'. While it is out of the scope of this analysis, it is relevant to note that the number of matches was significantly higher in the No Visual Feedback condition ($M=4.333$, $SD=1.784$) compared to the Visual Feedback condition ($M=2.14$, $SD=1.953$) ($F_{(1,22)}=9.354$, $p=0.006$, $\eta^2=0.263$, $d=1.17$).

Second, a chi-square analysis was performed to clarify the link between gender and *lexical innovation*. It was also necessary to address whether strategies and behaviours change when

communication problems, like errors and non-understandings, occur. As such, the variables tested were the gender of the instructor, occurrence of miscommunication, and the number of new words present in the next turn. In particular, this analysis considered the number of novel words contained in an utterance immediately after a (i) non-problematic and (ii) problematic utterance (that is, a dialogue utterance marked as a non-understanding, an incorrect instruction or in which an execution error occurred; a combined measure was used since the nature and cause of miscommunication was not the focus of this analysis). All utterances were grouped based on whether or not they contained novel words, and whether or not they followed a problematic utterance. The relationship between the variables was only found to be significant in the No Visual Feedback condition. Thus, the No

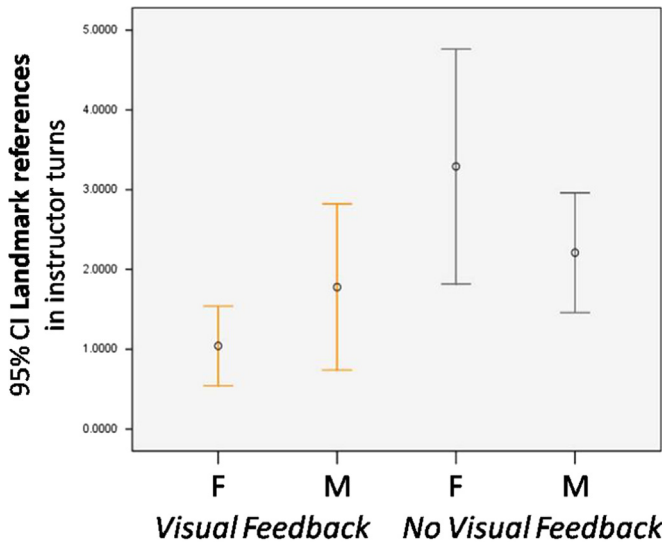


Fig. 7. References to landmarks in turns by female and male instructors in the Visual Feedback and No Visual Feedback conditions. Pronounced differences were found between female instructors in the two conditions (illustrated by the first and third error bars).

Visual Feedback data were used in this analysis to explore the gender factor.

A significant relationship was found between gender and the number of new words category after problematic and non-problematic utterances. After problem-free communication, both female and male instructors tended to re-use old vocabulary. 70% of utterances (for female instructors) and 64% (for males) comprised exclusively previously-used vocabulary. The Pearson's chi-square on the relationship yielded $\chi^2=8.035$, confirmed by the linear relationship of 8.031 ($df=1, p=0.005, \phi=0.063$). However, there was a difference between male and female instructors. The odds ratio of adhering to the old vocabulary was 1.3 higher for female instructors than for male instructors after a non-problematic utterance.

However, after miscommunication the analysis revealed that female and male instructors employed contrasting approaches. Male instructors responded by introducing new words (66% of the utterances). In contrast, female instructors appeared to continue adhering to the old vocabulary. The linear and Pearson's chi-square of 4.779 and 4.723, respectively, supported the existence of the relationship, statistically significant at $p=0.029$ (the phi coefficient was 0.233, explaining 5.42% of the variance). The odds ratio was 2.6, indicating that female instructors were now 2.6 times less likely to try new words after miscommunication

compared to males. In brief, the analysis suggests that females preferred to re-use vocabulary, even after miscommunication, whereas males were more inclined to introduce new words. The results are summarised in the graph in Fig. 9.

Pearson's chi-square analysis on similar experimental dialogue data appears to be common practice in related literature. However, strictly speaking, the use of Pearson's chi-square is incorrect, since the utterances were not independent from each other, being produced by the same 32 pairs. As such, this analysis also included the Cochran–Mantel–Haenszel (CMH) test, which has been proposed as an alternative method that can strengthen the reliability of the chi-square (Cochran, 1954). This test allows control for one variable (in this case, pair), while comparing the levels of the other two variables (in this case (i) miscommunication/no miscommunication and (ii) 1 or more new words/no new words). The

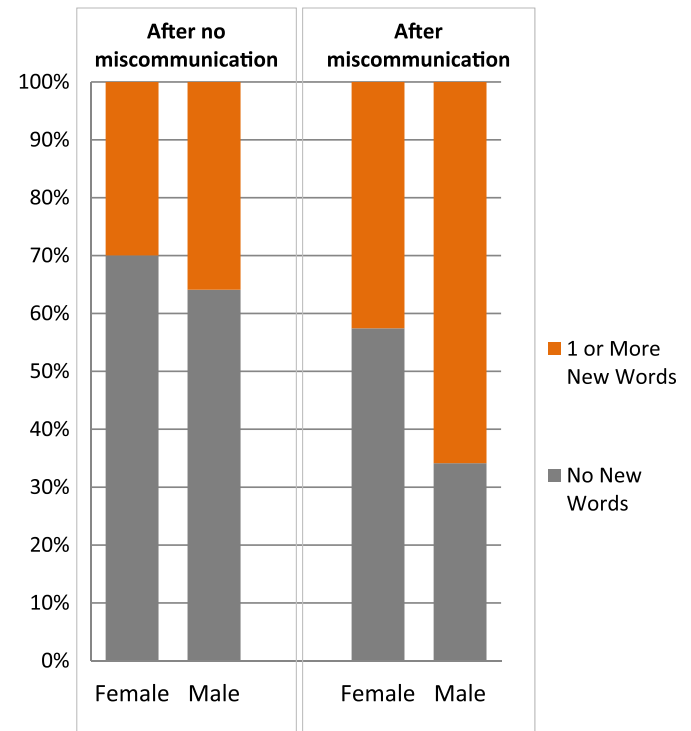


Fig. 9. Probability of occurrence of new words after problematic and non-problematic utterances for pairs with female and male instructors in the No Visual Feedback condition. Males try new words when miscommunication is detected, while females are more likely to re-use vocabulary.

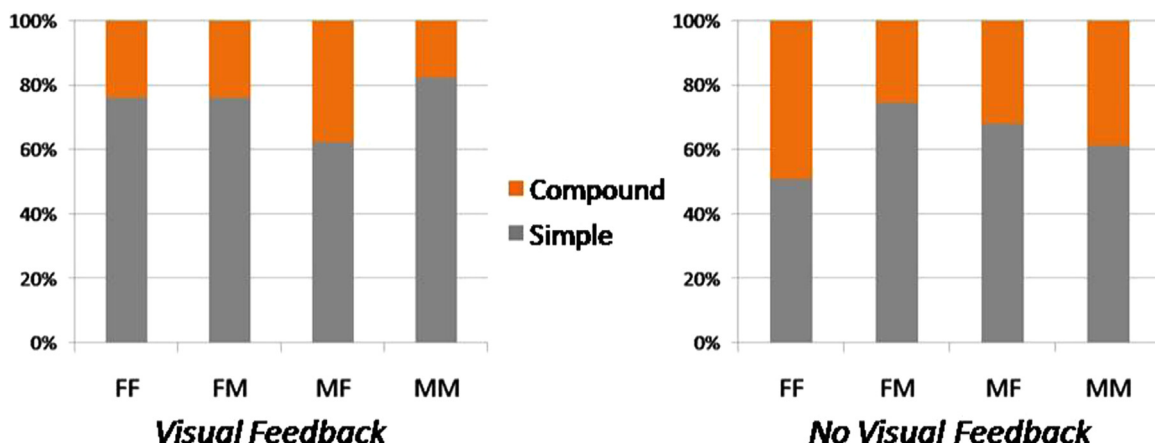


Fig. 8. The proportion of simple and compound instructions by instructors in all pair compositions in the Visual Feedback and No Visual Feedback conditions. Pronounced differences are found between FF pairs in the two conditions.

results of the CMH test confirmed the chi-square interpretation. It was verified that male instructors tend to produce considerably more utterances containing new words after miscommunication ($\chi^2=15.203$, $df=1$, $p=0.001$). The CMH test did not find a significant association between miscommunication and new words for pairs with female instructors, indicating that females tend to use previous vocabulary in both situations of problematic and problem-free communication.

Based on this analysis, there is insufficient evidence to support the suggestion that the tendency to converge linguistically in task-oriented dialogues depends on one's gender; as such, hypothesis H_{G5} is not validated. However, when communication problems occurred, gender differences were observed, with females tending to resort to previously used vocabulary and males introducing different words.

4.4. Summary of findings

In terms of performance, the gender factor hypothesis was not supported. Consistent with previous research, a broad benefit of visual feedback in terms of language economy and efficiency was found. However, there was an accuracy trade-off, as indicated by the higher frequency of incorrect instructions and non-understandings. While there was no difference in actual performance, females reported lower perceived task success than males. In terms of communication strategies, males were more likely to linguistically diverge and introduce novel words, especially in the event of communication problems. In the absence of visual feedback, females adapted their communication strategies significantly and effectively, increasing the quality and specificity of their verbal contributions, requesting and negotiating information, while the behaviour of males did not vary across conditions. Table 5 lists the 10 research hypotheses tested in the study and summarises the respective outcomes.

5. Discussion

This section discusses the practical implications of the findings and uses them to frame specific design recommendations for CSCW systems.

5.1. The role of visual feedback in CSCW

Awareness of how visual feedback affects collaboration and communication patterns can enable interface designers to take full advantage of its benefits and avoid related pitfalls. The analysis confirmed that visual feedback has a significant effect on CSCW, and revealed positive as well as negative dimensions of the effect, which will be discussed in this section.

5.1.1. Visual feedback leads to more efficient and economical CSCW interactions

The study confirmed the benefits of visually-supported collaborations by showing that sharing the workspace facilitates situation awareness and grounding. Visual feedback enables users of CSCW systems to complete their tasks more efficiently, with shorter interactions and simpler language. This phenomenon is illustrated by juxtaposing two dialogue excerpts from the Visual Feedback and No Visual Feedback conditions (shown in Table 6). It was also corroborated that when visual information is shared, much of the communication is carried out through physical actions, which replace verbal turns. As in the example from the Visual Feedback condition in Table 6, verbal communication by the follower appears redundant, and it was ignored by the instructor.

The results of this study confirmed that, without visual feedback, instructor queries increased, and followers needed to provide elaborate responses and descriptions, such that the responsibility for task and understanding maintenance was equally distributed. CSCW frequently involves remote training or help-giving dialogues between novices and experts (for example, Fussell et al., 2000; Karsenty, 1999; Dix, 1994; Crabtree et al., 2006; Twidale and Ruhleder, 2004) and, by definition, novices are unable to provide equal or precise contributions. For such applications, in which there is an expected prior asymmetry in the knowledge and

Table 5
List of research hypotheses and respective outcomes.

Hypotheses	Results and Commentary
H_{C1} : Gender has an effect on performance.	Not supported; while same-gender pairs were more accurate when giving instructions, no differences were found for execution errors and non-understandings.
H_{C2} : Gender has an effect on user perceptions.	Confirmed; females reported lower perceived task success than males.
H_{C3} : Gender has an effect on communication structure, in terms of frequency of queries, acknowledgements, and clarifications.	Confirmed; female instructors in all-female pairs in the No Visual Feedback condition provided the largest number of queries and acknowledgements.
H_{C4} : Gender has an effect on communication content, in terms of language specificity.	Confirmed; female instructors in all-female pairs provided utterances with higher specificity (higher frequency of landmark references and compound instructions), but only when visual feedback was withheld.
H_{C5} : Gender has an effect on communication behaviour, in terms of linguistic adaptation (lexical alignment).	Not supported; no differences in 'matches'. However, especially after communication problems, lexical innovation was higher for males.
H_{V1} : Visual feedback benefits performance.	Not supported; while fewer words were required, miscommunication increased when visual feedback was available.
H_{V2} : Visual feedback enables pairs to complete the task using lower frequency of queries, acknowledgements, and clarifications.	Confirmed; instructors issued fewer queries, and followers provided fewer acknowledgements and clarifications, when visual feedback was available.
H_{V3} : Visual feedback will enable pairs to complete the task with lower language specificity.	Confirmed; instructors omitted landmark references and action boundary information and provided simple instructions; followers omitted landmark references and did not state frame of reference in the Visual Feedback condition.
H_{CV1} : Visual feedback moderates the performance of males and females.	Not supported; no interaction effect was found for the performance variables.
H_{CV2} : Visual feedback moderates the communication strategies of males and females.	Confirmed; the interaction effects found for most communication variables suggest that all-female pairs fully adapted their strategies in the absence of visual feedback.

Table 6
Dialogue examples from the Visual Feedback (left-hand side) and No Visual Feedback (right-hand side) conditions.

Visual Feedback condition [MF8_S66-78]	No Visual Feedback condition [NMF6_F47-57]
I: go to the shop [movement]	I: go to the bridge and turn left before the bridge
I: move forwards [movement]	F: Done
F: is this right? [movement]	I: is the gas station on your left?
I: stop	F: Yes it is
I: turn around and then move forwards [movement]	I: now keep going down the road until you see a car park
F: do I go the other way? [movement]	F: I am in front of the car park
I: take the road on the right [movement]	I: turn right and walk till the end, along the road you will see a gym on your right
I: stop	F: Yes, gym to my right side
I: move forwards a little bit [movement]	I: good, keep going straight and you will see a factory on your left
F: am I here yet?	F: Yes, factory to my left side
I: move forwards [movement]	I: well done, goodbye
I: stop	
I: you're at your destination, goodbye	

proficiency states of the collaborators, the ability to ground information should be reinforced. Developers should consider implementing visual functionality to support grounding that does not rely on language, such as remote pointers, highlighting tools, and other methods that ensure joint attention to objects (Gergle, 2006; Fussell et al., 2003a, 2003b). Through such mechanisms, the expert can easily refer, and draw attention, to salient elements and details of the context when working with the novice who may be unable to comprehend or utilise domain-specific language. The results of this study also showed that when instructors can view the partner's workspace, the partner's actions function as verbal statements. As such, the novice may be able to use tools such as pointer trajectories (Gutwin and Penner, 2002; Fraser et al., 2007), which show the movement of the cursor, as a way to provide non-verbal feedback. Along the same lines, visually enriched interfaces can support non-native speakers of English, for whom linguistic aids may present an additional hurdle (as exemplified by a navigation study by Veinott et al. (1999)).

Clark and Brennan's (1991) framework supports the view that different mediums impose different costs on how people ground information. Speech is ephemeral, so people engage in a frequent grounding process of small chunks of language. In contrast, typed communication involves higher production costs, so interlocutors ground less frequently and through longer utterances. The results of this study also suggested that visual feedback improves communication efficiency. Since CSCW often relies on synchronous text-based communication (instant messaging), visual information is argued to be a primary requirement for such systems because it can alleviate some of the higher costs of grounding.

5.1.2. Visual feedback has a negative effect on accuracy

The results of this study also demonstrated the potential pitfalls by revealing a rise in miscommunication (non-understandings and incorrect executions) when partners shared visual information. This empirically supports the argument that visually-enhanced interactions may present a close, but misleading approximation to face-to-face communication, giving rise to misplaced assumptions of continuous joint perspective and common ground, and leading interlocutors to relax their grounding criteria. Therefore, although counter-intuitive, it is argued that when the available evidence of understanding is less solid and reliable (that is, only language, no visual feedback), the criteria to ensure that understanding is being

achieved become stricter, forcing interlocutors to be more accurate, persistent and detailed (and, consequently, less efficient in terms of word and turn usage, as also observed). In contrast, visual feedback relaxes the criteria and causes interlocutors to be less precise which, in turn, results in higher miscommunication. This phenomenon is further explained below in light of related literature and examples from the corpus.

The Collaborative Model postulates that interlocutors do not seek perfect and complete mutual understanding, but rather to sufficiently understand each other for current interaction purposes, meaning that grounding criteria are only as precise as they need to be (Brennan, 2005). In Brennan's study, followers collaborating in a spatial task reached the target more closely when instructors did not receive visual feedback. Two complete dialogues between pairs from the Visual Feedback and No Visual Feedback conditions from this study, provided in Table 7, illustrate this tendency. The destination in this case was the Lab. The instructor in the No Visual Feedback condition required that the follower not only reached but also went inside the location before asserting that the task was accomplished, whereas the instructor in the Visual Feedback condition provided directions that led the follower about 100 pixels off the target and ended the task (see image in Fig. 10). It is also interesting to note that instructors in the Visual Feedback condition did not usually state that this building was the destination, as in the dialogue example in Table 7. This confirms that visual feedback leads to inflated assumptions of what is mutually known or perceived. Finally, similar to the findings presented in this study, Brennan (2005) also observed that execution error rates were no higher without visual feedback than with it.

People try to complete a task by expending the least effort possible to achieve a satisfactory result; as Carletta and Mellish (1996, p.71) maintain, "in task-oriented dialogue, this produces a tension between conveying information carefully to the partner and leaving it to be inferred, risking a misunderstanding and the need for recovery." Thus, when the interaction conditions are deemed favourable (as in case of sharing visual feedback), speakers typically prefer to use expressions that may be more economical, but also increase ambiguity and the risk of incorrect interpretation. The conclusion that can be drawn is that more or stronger evidence does not necessarily lead to more successful interactions, but successful interactions depend on how well people are able to tune their grounding criteria.

This study showed a broad communication efficiency benefit, associated with an accuracy trade-off, and it has discussed practical implications for novice-expert and text-based dialogues. While it may be argued that these observations are linked to the properties of the visual feedback provided or the nature of the

Table 7
Dialogue examples from the Visual Feedback (left-hand side) and No Visual Feedback conditions (right-hand side).

Visual Feedback condition [MF9_L36-47]	No Visual Feedback condition [NMF8_L18-29]
I:hello	I: hello
F: hello	F: hello
I: go straight ahead	I: walk straight then turn right
I: turn right	F: now where do I go?
I: now, turn left then right	I: where are you now?
F: I have reached the junction	F: the pub is on my right
F: ok	I: walk straight past the pub and stop at the lab
F: straight ahead or turn left?	F: I am at the lab now
I: keep going straight	I: go into the lab
I: goodbye	F: I am inside the lab now
F: I have reached the junction by the bridge	I: goodbye
F: goodbye	F: goodbye



Fig. 10. The execution of the instructions provided in the dialogues in the table above (Table 7). The thick yellow line represents the path taken by both followers. The red dashed line and blue solid line show the finishing execution of the followers in the Visual Feedback condition and No Visual Feedback conditions, respectively.

task, they highlight the necessity to gain better awareness of how visual elements modify behaviours, interact with verbal communication, and integrate with it. For example, it remains unknown whether visual feedback impairs long-term task performance, especially in expert-novice collaborations, as suggested in Yuviler-Gavish et al. (2011), which showed that monitoring inhibits deep learning and exploration. Therefore, further systematic research should focus on identifying the benefits and pitfalls of visual feedback in CSCW.

5.2. The role of gender in CSCW

The analysis confirmed that the gender of the collaborators has a significant effect on CSCW. While the actual performance of males and females was comparable, there were differences in perceived performance. Differences were also observed in communication strategies, particularly in the absence of visual feedback, and when miscommunication occurred.

5.2.1. No actual performance difference, but lower perceived performance among females

This study supports the theoretical argument presented in Hyde (2005) that the impact of gender can be mitigated in interactive settings. Related empirical evidence can be found in other domains: studies in computer science education have reported that pair programming reduces the gender performance gap

between male and female programmers, and failure rates for students of both genders (Berenson et al., 2004; McDowell et al., 2003). The results of the study reported in this paper should be viewed in light of its task domain. Males perform consistently better in navigation tasks, particularly in virtual environments (VEs) (see Coluccia and Louse (2004), and Martens and Antonenko (2012)). In the navigation task employed in this study, the performance of females was comparable to the performance of males. As such, it may be deduced that the interactive CSCW setting enables female users to tackle difficulties associated with the nature of the task. This argument may be relevant for the design of VEs, given that they are widely used as training tools in several professional fields.

While task performance was comparable, a difference in perceptions was observed. It is a recurring research finding that females perceive their ability or performance in computer-based tasks to be lower than it actually is (Busch, 1995, 1996; Hargittai and Shafer, 2006). Female users also tend to attribute problems to their own lack of skill and are less likely to 'blame' the computer system (Beckwith and Burnett, 2004; Boiano et al., 2006), which may also explain the finding of this study that females gave higher ratings to the system than did males. The observation that actual performance may remain unaffected by these negative emotional states does not reduce the urgency of the problem. Poor self-perception leads to disengagement with computer-based activities and unwillingness to adopt technology or more advanced applications (Hartzel, 2003; Beckwith et al., 2006b; Bao et al., 2013; Venkatesh and Morris, 2000), and has also been argued to contribute to the underrepresentation of females in STEM fields (Chipman, 2005; Kinsey et al., 2008; Sáinz and López-Sáez, 2010).

The argument that CSCW provides a general performance advantage through interaction has implications for the adoption of related technologies in education and the workplace. This benefit may also relate to the idea that 'social genders' in CSCW become less relevant. In fact, a preliminary study with university students has suggested that CSCW settings enable more effective, balanced and less 'stereotyped' interactions between females and males than typical face-to-face classroom settings (Tomai et al., 2014). Yet, even if the effect of social biases associated with gender is alleviated, low self-efficacy among females also emerges in the domain of CSCW, which could compromise the acceptance of this technology.

5.2.2. All-female pairs compensate for the lack of visual cues through rich verbal means

As discussed in the previous subsection, the argument that interactive situations moderate gender differences performance is not new. However, there has been no focused attempt to pinpoint which elements of the communication underlie this effect. The results of this study help to outline these elements and validate the hypothesis by Savicki et al. (2006) and Prinsen et al. (2009) that females will overcome the paucity of CMC settings by deploying linguistic tools. In the present study, all-female pairs managed to compensate for the 'cueless' interaction condition by increasing the specificity of their contributions, and performed the task through elaborate, detailed verbal contributions and grounding information. Female partners working in pairs exhibited strong collaborative and adaptive behaviour, by putting in more communicative effort, successfully reducing uncertainty and attending to their partner, when the interaction conditions were poorer.

A study by Devlin and Bernstein (1995) serves to exemplify how this insight can motivate a design decision; in their study, females performed better, and equally well as males, when given the opportunity to complement the visual and map aid configuration with verbal instructions. Similarly, as Hubona and Shirah (2004) maintain, interfaces should be rich in static and dynamic

visual cues, but it should be possible to replace or complement some visually-presented information with verbal/textual content. Through such gender-neutral interfaces that offer the possibility of customisable settings (in this case, the provision of verbal aids), the needs and preferences of both genders can be met.

In seeking to understand the notable finding in relation to the performance of all-female pairs, it may be that the nature of the convergence in relation to the dialogue is important. Empirical models of human communication equate successful communication with convergence in the situation models of interlocutors – convergence is progressively reached, as the dialogue unfolds, through alignment (or adaptation) across all linguistic levels (Pickering and Garrod, 2004). If this proposition is considered in conjunction with the finding that males and females have gender-specific discourse styles in CMC (Herring and Stoerger, 2014), it may be argued that female-female pairs, *having the 'same' linguistic models to begin with*, were able to achieve convergence more quickly and more efficiently than mixed-gender pairs, who may have started with a weaker common ground, within the relatively short time that they had to interact with each other. However, the present study failed to produce consistent evidence in support of this interpretation, and, as such, it remains a conjecture to be targeted and further explored in future research.

5.2.3. Females use conservative strategies, while males engage in explorative behaviour

The study yielded unique findings in relation to what females and males do when faced with communication breakdowns. The analysis on lexical innovation demonstrated that females draw on previously used vocabulary, while males introduce new terms. Even in smooth communication, females are less willing to 'experiment' with novel expressions compared to male users. This may suggest that females are more conservative and males more explorative when handling communication breakdowns.

Among many possible explanations, this tendency could relate to gender differences in risk and cost perceptions. In the HCI domain, risk/cost perception is associated with users being less willing to try a useful but unfamiliar feature. Previous research argues that females perceive higher risks when they are involved in decisions or situations (for example, Finucane et al., 2000; Blais and Weber, 2001). As such, it is argued that females will be less likely to explore and experiment with unfamiliar features compared to males. Studies in various application domains, from programming IDEs (Beckwith et al., 2006a; Burnett et al., 2010; Cao et al., 2010) and spreadsheet software tools (Burnett et al., 2011) to web-based databases (Rosson et al., 2008) have confirmed that females are less confident to use novel software features while men typically engage in exploratory behaviour. It is argued that females' tendency to reuse vocabulary and not attempt a new strategy, even when these messages ostensibly failed, forms part of their general fear of 'tinkering' (the fear of trying new features), which may be traced back to females' low confidence and self-efficacy (Beckwith et al., 2006b) (as also discussed in the previous subsection).

Such findings should be considered by interface developers when unfamiliar or new features and strategies have to be adopted in the interaction with a system. In such situations, making use of techniques such as tutorial snippets, examples of what to say/do and short strategy explanations may help some users to feel more comfortable. In a gender-neutral interface, such features should be customisable in order to avoid compromising the experience of a gender group. Certainly, inclusive design does not mean that the experience of male users should be impaired – functionalities should be made optional. Since gender is a stable user profile characteristic, such options can be easily implemented in an adaptive system.

However, it should be emphasised that neither females nor males are homogeneous groups of users exhibiting all the characteristics

and preferences that are statistically associated with their gender. It is highly likely that many males are affected by the same interface complexities as females, and many females may enjoy the same software features as males (Beckwith et al., 2006a). This underscores the importance of gender-neutral software that supports all users. An interesting idea, proposed by Ljungblad and Holmquist (2007), is that designs that are informed by the needs and activities of a specific user population may also benefit the wider user population. For instance, verbal aids can provide support to users with a field-dependent cognitive style (Magoulas et al., 2004), as well as prove essential to users with visual impairments. Similarly, as argued in the previous subsection, verbal aids can catalyse the use and adoption of new software features by female users.

Finally, for many application domains, explorative and innovative user behaviour is not desirable. A good example is the (related to this study) domain of natural language interfaces, for which innovative and unpredictable user input is the main source of system failures.

6. Limitations and future work

Reflection on the methodology and results of this research has led to the identification of a number of limitations which shape directions for future experimental investigations, particularly, in the effect of gender dynamics in CSCW.

Section 3 discussed the effect of social aspects of gender, which motivated the decision to mask the gender of participants through a simulated human-robot interaction setup. As mentioned in Section 3.1.4, this experimental manipulation may limit the generalisability of the results because it does not directly map to a realistic CSCW setting, where information about the gender of the collaborators is available. In addition, the experimental manipulation created a knowledge asymmetry between participants, because only followers were aware that they were interacting with a person. In order to address these issues, the study should be replicated in two variations: a condition in which both interlocutors know they are collaborating with a person; and a condition in which interlocutors are told the gender of their partner. The comparative analysis of the results of the present study and the two variations could produce a comprehensive measure of the extent of the social effect of gender.

Second, the questionnaire used in this study was designed to be relatively short and simple in order for it to be completed after each task, and only the instructors' perceptions were captured. Therefore, the resulting observations are incomplete. A more sophisticated questionnaire tool tailored to the CSCW domain should be employed in a continuation study; for example, Convertino et al. (2007) have developed and validated a post-task questionnaire which specifically measures aspects of common ground, awareness, performance, interaction quality and satisfaction. The questionnaire should also target usability and affective factors, such as ease of learning, likeability, cognitive demands and annoyance, in order to provide insight into the role of gender (and gender interactions) in all dimensions of user experience.

The study focused on CSCW between pairs. A continuation of this work would investigate the effects of gender composition in group interactions. Indeed, initial empirical evidence suggests that mixed-gender groups report higher levels of satisfaction, social presence and performance (Wong et al. 2004; Houldsworth and Mathews, 2000; Hamlyn-Harris et al., 2006). Better understanding of the role of gender in group dynamics has important implications for the success of teamwork in organisational settings (Molyneux et al., 2008).

Several techniques were applied in order to reduce variance and increase internal validity, because of the relatively small sample size of this study (Sauro and Lewis, 2012, p. 121). However,

a homogeneous sample consisting of university students may affect the generalisability of the results. As such, a study with users with different demographic profiles needs to be undertaken. The characteristics that are likely to co-vary with gender, and are relevant to CSCW, include culture, age, education, task-related experience and computer expertise.

Another limitation of the study relates to the use of typed communication. The use of text-based dialogue enabled the experimental manipulation of masking the gender of participants. While the modality (speech or text) may not affect how spatial language is processed and formed (Tversky and Lee, 1999; Tversky et al., 2009), there are known differences between typed and spoken communication. In particular, typed communication is ‘quasi-synchronous’ (Garcia and Baker Jacobs, 1999); that is, the recipient sees the message in its entirety the moment his/her partner presses ‘enter’, whereas in spoken dialogue, interlocutors start formulating their response whilst listening to their partners’ utterance. This ‘quasi-synchrony’ may have also disrupted the sequential cohesion of dialogue, such that the second of two successive turns may not actually be the response to the first one (Herring, 1999). Moreover, as previously discussed, spoken dialogue involves more frequent grounding of shorter utterances. Finally, grounding is performed via auditory and gestural cues, while in text-based communication, mutual understanding is established through more explicit means. As such, further experimentation using speech or multi-modal interaction (for example, combining language-based communication with pointing or drawing) is needed to confirm, and complement, the findings of this study.

The methodology of this study, and the interpretation of the findings with respect to visual feedback, are grounded in previous empirical research. These studies have used a rich variety of collaborative tasks – for example, puzzle solving (Gergle et al., 2004, 2013), repairing (Kraut et al., 2003; Fussell et al., 2000), construction (Fussell et al., 2003a, 2003b; Clark and Krych, 2004) and navigation tasks (Anderson et al., 1991) – while also varying the level of task complexity (Ou et al., 2005). The diversification of experimental tasks is necessary, given that different tasks draw on different cognitive abilities and involve different collaboration activities and strategies. Therefore, in order to ensure that the gender effects revealed in this study were not merely a product of the properties and nature of the task, the investigation should be replicated using a variety of tasks drawn from these existing studies.

7. Conclusion

The study reported in this paper examined the effect of gender on CSCW, and explored whether visual feedback (often present in CSCW) moderates this effect. Two high-level conclusions can be drawn from this study: first, visual feedback results in more efficient but less accurate CSCW interactions; and, second, gender has a clear effect in terms of communication strategies and perceptions, but not in performance outcomes. It is hoped that the empirical contributions of this study will serve to stimulate further research in gender and human factors in CSCW. Unless design decisions are driven by research, systems are destined to include features that may be superfluous or even obstructive to particular groups of users. However, it is crucial that the objective of research in the user gender factor is to inform ‘gender-neutral’ systems; that is, the focus of research should shift from describing the differences towards describing interface features that are suitable for both genders.

Acknowledgments

Open access for this article was funded by King’s College London.

References

- Allen, G.L., 2000. Men and women, maps and minds: cognitive bases of sex-related differences in reading and interpreting maps. In: O’Nuallain, S. (Ed.), *Spatial Cognition: Foundations and Applications*. John Benjamins, Amsterdam, pp. 3–18.
- Anderson, A.H., Bader, M., Bard, E.G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., 1991. The HCRC map task corpus. *Lang. Speech* 34 (4), 351–366.
- Anderson, A.H., Smallwood, L., MacDonald, R., Mullin, J., Fleming, A., O’Malley, C., 2000. Video data and video links in mediated communication: what do users value? *Int. J. Hum.-Comput. Stud.* 52 (1), 165–187.
- Anderson, T., Sanford, A., Thomson, A., Ion, W., 2007. Computer-supported and face-to-face collaboration on design tasks. *Discourse Process.* 43 (3), 201–228.
- Bao, Y., Xiong, T., Hu, Z., Kibelloh, M., 2013. Exploring gender differences on general and specific computer self-efficacy in mobile learning adoption. *J. Educ. Comput. Res.* 491, 111–132.
- Bae, S., Lee, T., 2011. Gender differences in consumers’ perception of online consumer reviews. *Electronic Commerce Research*, 11(2), 201–214.
- Beckwith, L., 2007. *Gender HCI Issues in End-User Programming*, (Ph.D. Thesis, Oregon State University). Retrieved from (<https://ir.library.oregonstate.edu/xmlui/handle/1957/4954>).
- Beckwith, L., Burnett, M., 2004, September. Gender: An important factor in end-user programming environments? In *Visual Languages and Human Centric Computing*, 2004 IEEE Symposium on pp. 107–114.
- Beckwith, L., Burnett, M., Grigoreanu, V., Wiedenbeck, S., 2006a. HCI: What about the software? *Computer*, 83–87.
- Beckwith, L., Kissinger, C., Burnett, M., Wiedenbeck, S., Lawrance, J., Blackwell, A., Cook, C., 2006b, April. Tinkering and gender in end-user programmers’ debugging. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* pp. 231–240. ACM.
- Belkadi, F., Bonjour, E., Camargo, M., Troussier, N., Eynard, B., 2013. A situation model to support awareness in collaborative design. *Int. J. Hum.-Comput. Stud.* 711, 110–129.
- Berenson, S.B., Slaten, K.M., Williams, L., Ho, C.W., 2004. Voices of women in a software engineering course: reflections on collaboration. *J. Educ. Resour. Comput.* 41, 3.
- Bernard, M., Mills, M., Friend, C., 2000. Male and female attitudes towards computer-mediated group interactions. *Internet Publ. Usability News*, 22.
- Bilous, F.R., Krauss, R.M., 1988. Dominance and accommodation in the conversational behaviours of same-and mixed-gender dyads. *Lang. Commun.* 8 (3), 183–194.
- Bimber, B., 2000. Measuring the gender gap on the Internet. *Soc. Sci. Quarterly* 813, 868–876.
- Blackwell, A.F., Rode, J.A., Toye, E.F., 2009. How do we program the home? Gender, attention investment, and the psychology of programming at home. *Int. J. Hum.-Comput. Stud.* 674, 324–341.
- Blais, A.R., Weber, E.U., 2001. Domain-specificity and gender differences in decision making. *Risk Decis. Policy* 61, 47–69.
- Boiano, S., Borda, A., Bowen, J., Faulkner, X., Gaia, G., McDavid, S., 2006. Gender issues in HCI design for web access. *Advances in Universal Web Design and Evaluation: Research, Trends and Opportunities*, Section III, Gender Issues, 116–153.
- Brennan, S.E., 2005. How conversation is shaped by visual and spoken evidence. In: Trueswell, J.C., Tanenhaus, M.K. (Eds.), *Approaches to Studying World-situated Language Use: Bridging the Language-as-product and Language-as-action Traditions*. MIT Press, Cambridge, MA, pp. 95–129.
- Brennan, S., Clark, H.H., 1996. Conceptual pacts and lexical choice in conversation. *J. Exp. Psychol.: Learn. Mem. Cogn.* 22, 1482–1493.
- Brennan, S.E., Lockridge, C.B., 2006. Computer-mediated communication: a cognitive science approach. *Encycl. Lang. Linguist.*, 775–780.
- Bromme, R., Jucks, R., Runde, A., 2005. *Barriers and Biases in Computer-mediated Expert-layperson-Communication*. Springer, US, pp. 89–118.
- Burnett, M.M., Beckwith, L., Wiedenbeck, S., Fleming, S.D., Cao, J., Park, T.H., Rector, K., 2011. Gender pluralism in problem-solving software. *Interact. Comput.* 235, 450–460.
- Burnett, M., Fleming, S.D., Iqbal, S., Venolia, G., Rajaram, V., Farooq, U., ... Czerwinski, M., 2010, September. Gender differences and programming environments: across programming populations. In *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*. ACM.
- Busch, T., 1995. Gender differences in self-efficacy and attitudes toward computers. *J. Educ. Comput. Res.* 122, 147–158.
- Busch, T., 1996. Gender, group composition, cooperation, and self-efficacy in computer studies. *J. Educ. Comput. Res.* 15, 125–136.
- Cao, J., Riche, Y., Wiedenbeck, S., Burnett, M., Grigoreanu, V., 2010, April. End-user mash up programming: through the design lens. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 1009–1018.
- Carletta, J., Mellish, C.S., 1996. Risk-taking and recovery in task-oriented dialogue. *J. Pragmat* 261, 71–107.
- Carletta, J., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G., Anderson, A., 1996. *HCRC dialogue structure coding manual*. Universities of Edinburgh and Glasgow.
- Carr, T., Cox, L., Eden, N., Hanslo, M., 2004. From peripheral to full participation in a blended trade bargaining simulation. *Br. J. Educ. Technol.* 352, 197–211.
- Carroll, J.M., Neale, D.C., Isenhour, P.L., Rosson, M.B., McCrickard, D.S., 2003.

- Notification and awareness: synchronizing task-oriented collaborative activity. *Int. J. Hum.-Comput. Stud.* 585, 605–632.
- Carroll, J.M., Rosson, M.B., Convertino, G., Ganoë, C.H., 2006. Awareness and teamwork in computer-supported collaborations. *Interact. Comput.* 18(1), 21–46.
- Cassell, J., 1998. Chess for girls?: Feminism and computer games. In: Cassell, J., Jenkins, H. (Eds.), *From Barbie to Mortal Kombat: Gender and Computer Games*. MIT Press, Cambridge, MA.
- Chen, S.Y., Macredie, R., 2010. Web-based interaction: a review of three important human factors. *Int. J. Inform. Manag.* 30(5), 379–387.
- Chipman, S.F., 2005. Research on the Women and Mathematics Issue: A Personal Case History, Gender Differences in Mathematics. Ed. Ann M. Gallagher, James C. Kaufman. 1st ed. Cambridge: Cambridge University Press, 1–24. Cambridge Books Online.
- Choi, K.S., Deek, F.P., Im, I., 2009. Pair dynamics in team collaboration. *Comput. Hum. Behav.* 25(4), 844–852.
- Clark, H.H., 1996. *Using Language*. Cambridge University Press, New York, US.
- Clark, H.H., Brennan, S.E., 1991. Grounding in communication. *Perspect. Soc. Sci. Cogn.* 13, 127–149.
- Clark, H.H., Marshall, C.R., 1981. Definite reference and mutual knowledge. *Psycholinguist: Crit. Concepts Psychol.* 3, 414.
- Clark, H.H., Wilkes-Gibbs, D., 1986. Referring as a collaborative process. *Cognition* 22(1), 1–39.
- Clark, H.H., Krych, M.A., 2004. Speaking while monitoring addressees for understanding. *Mem. Lang. J.* 50, 62–81.
- Coates, J., 1986. *Women, Men and Languages: Studies in Language and Linguistics*. Longman, London.
- Cochran, W.G., 1954. Some methods for strengthening the common χ^2 tests. *Biometrics* 10(4), 417–451.
- Cockburn, A., McKenzie, B., 2002. Evaluating the effectiveness of spatial memory in 2D and 3D physical and virtual environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*. pp. 203–210. ACM.
- Coluccia, E., Louse, G., 2004. Gender differences in spatial orientation: a review. *J. Environ. Psychol.* 24(3), 329–340.
- Convertino, G., Mentis, H.M., Slavkovic, A., Rosson, M.B., Carroll, J.M., 2011. Supporting common ground and awareness in emergency management planning: a design research project. *ACM Trans. Comput.-Hum. Interact.* 18(4), 22.
- Convertino, G., Mentis, H.M., Ting, A.Y., Rosson, M.B. and Carroll, J.M., 2007, November. How does common ground increase? In *Proceedings of the 2007 international ACM conference on Supporting group work*. pp. 225–228. (ACM).
- Crabtree, A., O'Neill, J., Tolmie, P., Castellani, S., Colombino, T., Grasso, A., 2006, November. The practical indispensability of articulation work to immediate and remote help-giving. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work* pp. 219–228. ACM.
- Cutmore, T.R., Hine, T.J., Maberly, K.J., Langford, N.M., Hawgood, G., 2000. Cognitive and gender factors influencing navigation in a virtual environment. *Int. J. Hum.-Comput. Stud.* 53(2), 223–249.
- Daly-Jones, O., Monk, A., Watts, L., 1998. Some advantages of video conferencing over high-quality audio conferencing: fluency and awareness of attentional focus. *Int. J. Hum.-Comput. Stud.* 49(1), 21–58.
- Denis, M., 1997. The description of routes: a cognitive approach to the production of spatial discourse. *Cahiers Psychol. Cogn.* 16(4), 409–458.
- Devlin, A.S., Bernstein, J., 1995. Interactive wayfinding: use of cues by men and women. *J. Environ. Psychol.* 15(1), 23–38.
- Dillon, A., Watson, C., 1996. User analysis in HCI: the historical lesson from individual differences research. *Int. J. Hum.-Comput. Stud.* 45(6), 619–637.
- Ding, N., Bosker, R.J., Harskamp, E.G., 2011. Exploring gender and gender pairing in the knowledge elaboration processes of students using computer-supported collaborative learning. *Comput. Educ.* 56(2), 325–336.
- Dix, A., 1994. *Computer supported cooperative work: a framework*, Design Issues in CSCW. Springer, London, pp. 9–26.
- Djamasbi, S., Loiacono, E.T., 2008. Do men and women use feedback provided by their Decision Support Systems DSS differently? *Decis. Support Syst.* 44(4), 854–869.
- Dourish, P., Bellotti, V., 1992, December. Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work* pp. 107–114. ACM.
- Durndell, A., Haag, Z., Laithwaite, H., 2000. Computer self efficacy and gender: a cross cultural study of Scotland and Romania. *Personal. Individ. Differ.* 28(6), 1037–1044.
- Finucane, M.L., Slovic, P., Mertz, C.K., Flynn, J., Satterfield, T.A., 2000. Gender, race, and perceived risk: the 'white male' effect. *Health Risk Soc.* 22, 159–172.
- Fischer, A.H., 2011. Gendered social interactions in face-to-face and computer mediated communication. *Face-to-face communication over the Internet*, 53–78.
- Fitzpatrick, M.A., Mulac, A., Dindia, K., 1995. Gender-preferential language use in spouse and stranger interaction. *J. Lang. Soc. Psychol.* 14(1–14), 18–39.
- Fraser, M., McCarthy, M.R., Shaikat, M., Smith, P., 2007. Seconds matter: improving distributed coordination by tracking and visualizing display trajectories. In *CHI2007/Vol. 2*, p. 1303. ACM.
- Fussell, S.R., Kraut, R.E., Siegel, J., 2000, December. Coordination of communication: Effects of shared visual context on collaborative work. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work* pp. 21–30. ACM.
- Fussell, S.R., Setlock, L.D., 2014. *Computer-mediated communication*. In: Holtgraves, T. (Ed.), *Handbook of Language and Social Psychology*. Oxford University Press, Oxford, UK, pp. 471–490.
- Fussell, S.R., Setlock, L.D., Kraut, R.E., 2003a. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. *Proceedings of CHI 2003* pp. 513–520. NY: ACM Press.
- Fussell, S.R., Setlock, L.D., Parker, E.M., Yang, J., 2003, April. Assessing the value of a cursor pointing device for remote collaboration on physical tasks. In *CHI'03 Extended Abstracts on Human Factors in Computing Systems* pp. 788–789. ACM.
- Garcia, A.C., Baker Jacobs, J., 1999. The eyes of the beholder: Understanding the turn-taking system in quasi-synchronous computer-mediated communication. *Research on language and social interaction*, 32(4), 337–367.
- Garrod, S., Anderson, A., 1987. Saying what you mean in dialogue: a study in conceptual and semantic co-ordination. *Cognition* 27(2), 181–218.
- Garrod, S., Pickering, M.J., 2004. Why is conversation so easy? *Trends Cogn. Sci.* 8(1), 8–11.
- Gergle, D.R., 2006. *The value of shared visual information for task-oriented collaboration* PhD Thesis, Carnegie Mellon University. Pittsburgh, PA. Retrieved from: dgergle.soc.northwestern.edu/resources/Gergle_Dissertation2006.pdf.
- Gergle, D., Kraut, R.E., Fussell, S.R., 2004. Language efficiency and visual technology minimizing collaborative effort with visual information. *J. Lang. Soc. Psychol.* 23(4), 491–517.
- Gergle, D., Kraut, R.E., Fussell, S.R., 2013. Using visual information for grounding and awareness in collaborative tasks. *Hum.-Comput. Interact.* 28, 1.
- Giles, H., Coupland, N., Coupland, I., 1991. Accommodation theory: Communication, context, and. *Contexts of accommodation: Developments in applied sociolinguistics*, 1.
- Gladstone, E., O'Connor, K.M., 2014. A counterpart's feminine face signals cooperativeness and encourages negotiators to compete. *Organ Behav. Hum. Decis. Process.* 125(1), 18–25.
- Gutwin, C., Penner, R., 2002, November. Improving interpretation of remote gestures with telepointer traces. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work* pp. 49–57. ACM.
- Halpern, D.F., 2000. *Sex differences in cognitive abilities*. Mahwah, NJ: Lawrence Erlbaum.
- Halpern, D.F., Benbow, C.P., Geary, D.C., Gur, R.C., Hyde, J.S., Gernsbacher, M.A., 2007. The science of sex differences in science and mathematics. *Psychol. Sci. Public Interest* 8(1), 1–51.
- Hamlyn-Harris, J., Hurst, B., Von Baggo, K., Bayley, A., 2006. Predictors of team work satisfaction. *J. Inform. Technol. Educ.: Res.* 5(1), 299–315.
- Hargittai, E., Shafer, S., 2006. Differences in actual and perceived online skills: the role of gender. *Soc. Sci. Quarterly* 87(2), 432–448.
- Hartmann, T., Klimmt, C., 2006. Gender and computer games: exploring females' dislikes. *J. Comput.-Mediat. Commun.* 11(4), 910–931.
- Hartzel, K., 2003. How self-efficacy and gender issues affect software adoption and use. *Commun. ACM* 46(9), 167–171.
- Herring, S.C., 1999. Interactional coherence in CMC. *J. Comput. Mediat. Commun.* 4, 1.
- Herring, S.C., 2000. Gender differences in CMC: findings and implications. *Comput. Prof. Soc. Responsib. J.*, 181.
- Herring, S.C., Stoerger, S., 2014. Gender and (a)nonymity in computer-mediated communication. In: Ehrlich, S., Meyerhoff, M., Holmes, J. (Eds.), *The Handbook of Language, Gender, and Sexuality*, 2nd edition John Wiley Sons, Ltd, Chichester, UK, pp. 567–586.
- Holmes, J., 1990. Hedges and boosters in women's and men's speech. *Lang. Commun.* 10(3), 185–205.
- Houldsworth, C., Mathews, B.P., 2000. Group composition, performance and educational attainment. *Educ. Train.* 42(1), 40–53.
- Hubona, G.S., Shirah, G.W., 2004. The gender factor performing visualization tasks on computer media. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences* pp. (9-pp. IEEE).
- Hyde, J.S., 2005. The gender similarities hypothesis. *Am. Psychol.* 60(6), 581.
- Inzlicht, M., Ben-Zeev, T., 2000. A threatening intellectual environment: why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychol. Sci.* 11(5), 365–371.
- Isaacs, E.A., Clark, H.H., 1987. References in conversation between experts and novices. *J. Exp. Psychol.: Gen.* 116(1), 26.
- Karsenty, L., 1999. Cooperative work and shared visual context: an empirical study of comprehension problems in side-by-side and remote help dialogues. *Hum.-Comput. Interact.* 14(3), 283–315.
- Kinsey, B.L., Towle, E., O'Brien, E.J., Bauer, C.F., 2008. Analysis of self-efficacy and ability related to spatial tasks and the effect on retention for students in engineering. *Int. J. Eng. Educ.* 24(3), 488–494.
- Kling, K.C., Hyde, J.S., Showers, C.J., Buswell, B.N., 1999. Gender differences in self-esteem: a meta-analysis. *Psychol. Bull.* 125(4), 470.
- Klippel, A., Richter, K.F., Hansen, S., 2009. Cognitively ergonomic route directions. *Handbook of Research on Geoinformatics*, 230–238.
- Koh, Y.J., Sundar, S.S., 2010. Effects of specialization in computers, web sites, and web agents on e-commerce trust. *Int. J. Hum.-Comput. Stud.* 68(12), 899–912.
- Koulouri, T., Lauria, S., Macredie, R.D., Chen, S., 2012. Are we there yet?: the role of gender on the effectiveness and efficiency of user-robot communication in navigational tasks. *ACM Trans. Comput.-Hum. Interact.* 19(1), 4.
- Kraut R.E., Fussell S.R., Brennan S.E. and Siegel J., Understanding effects of proximity on collaboration: implications for technologies to support remote collaborative work. *Distrib. Work* 2002b, 137–162.
- Kraut, R.E., Fussell, S.R., Siegel, J., 2003. Visual information as a conversational resource in collaborative physical tasks. *Hum.-Comput. Interact.* 18(1), 13–49.
- Kraut, R.E., Gergle, D., Fussell, S.R., 2002a, November. The use of visual information

- in shared visual spaces: Informing the development of virtual co-presence. In Proceedings of the 2002 ACM conference on Computer supported cooperative work pp. 31–40. ACM.
- Lawton, C.A., 1994. Gender differences in way-finding strategies: relationship to spatial ability and spatial anxiety. *Sex Roles* 30 (11), 765–779.
- Levin, D.T., Killingsworth, S.S., Saylor, M.M., 2008. Concepts about the capabilities of computers and robots: A test of the scope of adults' theory of mind. In Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction (pp. 57–64). (ACM).
- Levinson, S.C., 1983. *Pragmatics*. Cambridge University, Cambridge, England.
- Lin, P.C., Chien, L.W., 2010. The effects of gender differences on operational performance and satisfaction with car navigation systems. *Int. J. Hum.-Comput. Stud.* 68(10), 777–787.
- Ljungblad, S., Holmquist, L.E., 2007, April. Transfer scenarios: grounding innovation with marginal practices. In Proceedings of the SIGCHI conference on Human factors in computing systems pp. 737–746. ACM.
- Lovelace, K., Hegarty, M., Montello, D., 1999. Elements of good route directions in familiar and unfamiliar environments. *Spatial information theory. Cognitive and computational foundations of geographic information science*, 751–751.
- Magoulas, G., Chen, S., Dimakopoulos, D., 2004. A personalised interface for web directories based on cognitive styles. *User-Centered Interaction Paradigms for Universal Access in the Information Society*, 159–166.
- Martens, J., Antonenko, P.D., 2012. Narrowing gender-based performance gaps in virtual environment navigation. *Comput. Hum. Behav.* 28(3), 809–819.
- Matheson, K., 1991. Social cues in computer-mediated negotiations: gender makes a difference. *Comput. Hum. Behav.* 7 (3), 137–145.
- McDowell, C., Werner, L., Bullock, H.E., Fernald, J., 2003, May. The impact of pair programming on student performance, perception and persistence. In Proceedings of the 25th international conference on Software engineering pp. 602–607. IEEE Computer Society.
- McRoy, S.W., 1998. Achieving robust human-computer communication. *Int. J. Hum.-Comput. Stud.* 48(5), 681–704.
- Mills, G.J., 2007. *Semantic Co-ordination in Dialogue: The Role of Direct Interaction* (PhD thesis). Queen Mary University, London, UK, Retrieved from (http://www.dcs.qmul.ac.uk/tech_reports/RR-07-07.pdf).
- Molyneux, H., Donnell, S.O., Fournier, H., Gibson, K., 2008, June. Participatory videoconferencing for groups. In *Technology and Society*, 2008. ISTAS 2008. IEEE International Symposium on pp. 1–8. (IEEE).
- Mulac, A., Wiemann, J.M., Widenmann, S.J., Gibson, T.W., 1988. Male/female language differences and effects in same-sex and mixed-sex dyads: the gender-linked language effect. *Commun. Monogr.* 55, 315–335.
- Namy, L.L., Nygaard, L.C., Sauerteig, D., 2002. Gender differences in vocal accommodation: the role of perception. *J. Lang. Soc. Psychol.* 21(4), 422–432.
- Norman, R., 1976. When what is said is important: a comparison of expert and attractive sources. *J. Exp. Soc. Psychol.* 12 (3), 294–300.
- O'Malley, C., Langton, S., Anderson, A., Doherty-Sneddon, G., Bruce, V., 1996. Comparison of face-to-face and video-mediated interaction. *Interact. Comput.* 8(2), 177–192.
- Ong, C.S., Lai, J.Y., 2006. Gender differences in perceptions and relationships among dominants of e-learning acceptance. *Comput. Hum. Behav.* 22(5), 816–829.
- Ou, J., Oh, L.M., Yang, J., Fussell, S.R., 2005, April. Effects of task properties, partner actions, and message content on eye gaze patterns in a collaborative task. In Proceedings of the SIGCHI conference on Human factors in computing systems. pp. 231–240. ACM.
- Patel, H., Pettitt, M., Wilson, J.R., 2012. Factors of collaborative working: a framework for a collaboration model. *Appl. Ergon.* 43(1), 1–26.
- Pickering, M.J., Garrod, S., 2004. The interactive-alignment model: developments and refinements. *Behav. Brain Sci.* 27(2), 212–225.
- Picucci, L., Caffò, A.O., Bosco, A., 2011. Besides navigation accuracy: gender differences in strategy selection and level of spatial confidence. *J. Environ. Psychol.* 31(4), 430–438.
- Prinsen, F.R., Volman, M.L.L., Terwel, J., 2007. Gender-related differences in computer-mediated communication and computer-supported collaborative learning. *J. Comput. Assist. Learn.* 23(5), 393–409.
- Prinsen, F.R., Volman, M.L.L., Terwel, J., Van den Eeden, P., 2009. Effects on participation of an experimental CSCL-programme to support elaboration: do all students benefit? *Comput. Educ.* 52(1), 113–125.
- Ranjan, A., Birnholtz, J.P., Balakrishnan, R., 2007, April. Dynamic shared visual spaces: experimenting with automatic camera control in a remote repair task. In Proceedings of the SIGCHI conference on Human factors in computing systems pp. 1177–1186. ACM.
- Richert, D., Halabi, A., Eaglin, A., Edwards, M., Bardzell, S., 2011, May. Arrange-A-Space: tabletop interfaces and gender collaboration. In Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems pp. 1495–1500. ACM.
- Rosson, M.B., Sinha, H., Bhattacharya, M., Zhao, D., 2008. Design planning by end-user web developers. *J. Vis. Lang. Comput.* 19(4), 468–484.
- Sanford, A., Anderson, A., Mullin, J., 2003. Adapting communicative strategies to computer mediated communication: an analysis of task performance and dialogue structure. *Applied Cognitive Psychology*, 17(3), 325–348.
- Sacks, H., Schegloff, E.A., Jefferson, G., 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 696–735.
- Sáinz, M., López-Sáez, M., 2010. Gender differences in computer attitudes and the choice of technology-related occupations in a sample of secondary students in Spain. *Comput. Educ.* 54(2), 578–587.
- Sandstrom, N.J., Kaufman, J., Huettel, S. A., 1998. Males and females use different distal cues in a virtual environment navigation task. *Cogn. Brain Res.* 64, 351–360.
- Saucier, D.M., Green, S.M., Leason, J., MacFadden, A., Bell, S., Elias, L.J., 2002. Are sex differences in navigation caused by sexually dimorphic strategies or by differences in the ability to use the strategies? *Behav. Neurosci.* 116(3), 403.
- Sauro, J., Lewis, J.R., 2012. *Quantifying the User Experience: Practical Statistics for User Research*. Morgan Kaufmann.
- Savicki, V., Foster, D.A., Kelley, M., 2006. Gender, Group Composition and Task Type in Virtual Groups. *Gender and communication at work*, 270.
- Savicki, V., Kelley, M., 2000. Computer mediated communication: gender and group composition. *CyberPsychol. Behav.* 35, 817–826.
- Savicki, V., Kelley, M., Lingenfelter, D., 1996. Gender and group composition in small task groups using computer-mediated communication. *Comput. Hum. Behav.* 12(2), 209–224.
- Schmader, T., Johns, M., Forbes, C., 2008. An integrated process model of stereotype threat effects on performance. *Psychol. Rev.* 115(2), 336.
- Schober, M.F., 1993. Spatial perspective-taking in conversation. *Cognition* 47(1), 1–24.
- Schober, M.F., 2009. *Spatial dialogue between partners with mismatched abilities*. In: Coventry, K.R., Tenbrink, T., Bateman, J. (Eds.), *Spatial Language and Dialogue*. Oxford University Press, Oxford, pp. 23–39.
- Sellen, A.J., 1995. Remote conversations: the effects of mediating talk with technology. *Hum.-Comput. Interact.* 10 (4), 401–444.
- Setlock, L.D., Fussell, S.R. and Neuwirth, C., 2004, November. Taking it out of context: collaborating within and across cultures in face-to-face settings and via instant messaging. In Proceedings of the 2004 ACM conference on Computer supported cooperative work. pp. 604–613. ACM.
- Setlock, L.D., Quinones, P.A., Fussell, S.R., 2007, January. Does culture interact with media richness? The effects of audio vs. video conferencing on Chinese and American dyads. In *System Sciences*, 2007. HICSS 2007. 40th Annual Hawaii International Conference on. pp. 13–13.
- Shih, M., Ambady, N., Richeson, J.A., Fujita, K., Gray, H.M., 2002. Stereotype performance boosts: the impact of self-relevance and the manner of stereotype activation. *J. Personal. Soc. Psychol.* 83(3), 638.
- Stein, R., Brennan, S.E., 2004, October. Another person's eye gaze as a cue in solving programming problems. In Proceedings of the 6th international conference on Multimodal interfaces (pp. 9–15). ACM.
- Stefik, M., Foster, G., Bobrow, D.G., Kahn, K., Lanning, S., Suchman, L., 1987. Beyond the chalkboard: computer support for collaboration and problem solving in meetings. *Commun. ACM* 30(1), 32–47.
- Mazei, J., Hüffmeier, J., Freund, P.A., Stuhlmacher, A.F., Bilke, L., Hertel, G., 2015. A meta-analysis on gender differences in negotiation outcomes and their moderators. *Psychol. Bull.* 141 (1), 85.
- Stupka, R., 2011. *Communication Accommodation in Mixed Gender Dyads*. Oshkosh Scholar. Vol. 6.
- Sun, X., 2008. *Why gender matters in CMC: gender differences in remote trust and performance with initial social activities* Ph.D. Thesis, Drexel University, Philadelphia, PA, Retrieved from <<http://idea.library.drexel.edu/handle/1860/2930>> .
- Tannen, D., 1989. *Talking voices: Repetition, dialogue and imagery in conversational discourse* Vol. 6. Cambridge, England: Cambridge University Press.
- Tenbrink, T., Hui, S., 2007. Negotiating spatial goals with a wheelchair. Proceedings of the 8th SIGDial.
- Tenbrink, T., Ross, R.J., Thomas, K.E., Dethlefs, N., Andonova, E., 2010. Route instructions in map-based human-human and human-computer dialogue: a comparative analysis. *J. Vis. Lang. Comput.* 21(5), 292–309.
- Tomai, M., Mebane, M.E., Rosa, V., Benedetti, M., 2014. Can computer supported collaborative learning CSCL promote counter-stereotypical gender communication styles in male and female university students? *Proc.-Soc. Behav. Sci.* 116, 4384–4392.
- Tversky, B., Lee, P., 1999. Pictorial and verbal tools for conveying routes. *Spatial information theory. Cogn. Comput. Found Geogr. Inform. Sci.*, 752.
- Tversky, B., Heiser, J., Lee, P., Daniel, M.P., 2009. Explanations in gesture, diagram, and word. In: Coventry, K.R., Tenbrink, T., Bateman, J. (Eds.), *Spatial Language and Dialogue*. Oxford University Press, Oxford.
- Twidale, M., Ruhlleder, K., 2004, November. Where am I and who am I?: issues in collaborative technical help. In Proceedings of the 2004 ACM conference on Computer supported cooperative work pp. 378–387. ACM.
- Ullman, M.T., Miranda, R.A., Travers, M.L. 2008. Sex differences in the neurocognition of language. *Sex differences in the brain: From genes to behavior*, 291–310.
- Underwood, J., Underwood, G., Wood, D., 2000. When does gender matter?: Interactions during computer-based problem solving. *Learn. Instr.* 10(5), 447–462.
- Vanetti, E.J., Allen, G.L., 1988. Communicating environmental knowledge the impact of verbal and spatial abilities on the production and comprehension of route directions. *Environ. Behav.* 20(6), 667–682.
- Veinott, E.S., Olson, J., Olson, G.M., Fu, X., 1999. Video helps remote work: Speakers who need to negotiate common ground benefit from seeing each other. In Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit pp. 302–309. ACM.
- Venkatesh, V., Morris, M., 2000. Why don't men ever stop to ask for direction? Gender, social influence, and their role in technology acceptance and usage behavior. *MIS Quarterly* 24(1), 115–139.
- Whittaker, S., 2003a. Things to talk about when talking about things. *Hum.-Comput. Interact.* 18(1–182), 149–170.
- Whittaker, S., 2003b. Theories and methods in mediated communication. The

- handbook of discourse processes, 243–286.
- Whittaker, S., Geelhoed, E., Robinson, E., 1993. Shared workspaces: how do they work and when are they useful? *Int. J. Man-Mach. Stud.* 395, 813–842.
- Wong, Y.K., Shi, Y., Wilson, D., 2004. Experience, gender composition, social presence, decision process satisfaction and group performance, ACM International Conference Proceeding Series, 58, 1–10.
- Wu, A., Convertino, G., Ganoë, C., Carroll, J.M., Zhang, X.L., 2013. Supporting collaborative sense-making in emergency management through geo-visualization. *Int. J. Hum.-Comput. Stud.* 71 (1), 4–23.
- Yoon, S.Y., Choi, Y.J., Oh, H., 2015. User attributes in processing 3D VR-enabled showroom: gender, visual cognitive styles, and the sense of presence. *Int. J. Hum.-Comput. Stud.*
- Yuviler-Gavish, N., Yechiam, E., Kallai, A., 2011. Learning in multimodal training: visual guidance can be both appealing and disadvantageous in spatial tasks. *Int. J. Hum.-Comput. Stud.* 693, 113–122.
- Zimbardo, P.G., Leippe, M.R., 1991. The psychology of attitude change and social influence. New York, NY: Mcgraw-Hill Book Company.



Rob Macredie is Professor of Interactive Systems in the Department of Computer Science, Brunel University London. Rob's key interest lies in the way in which people and organisations use technology, and his research aims to determine how work can be more effectively undertaken by improving the way that we understand how people and technology interact in organisational (and social) settings. He holds a BSc (hons) in Physics and Computer Science and a PhD in Computer Science and is a Fellow of the British Computer Society. He is Editor-in-Chief of *Virtual Reality* and an Editor of *Personal and Ubiquitous Computing*.



Theodora Koulouri is a Teaching Fellow in the Department of Informatics, King's College London. Theodora's research interests lie in the design and evaluation of systems that can support more efficient and natural interactions. She received her PhD from Brunel University London in 2013, and MSc degree from University College London.



Stanislao Lauria has a Laurea in Physics and Cybernetics from Università Federico II, Napoli (Italy) and a PhD in Cybernetics from the University of Reading (UK). He is a Lecturer at Brunel University (UK). He has been research fellow at both the University of Reading and University of Plymouth (UK). He is a computer scientist with an interest in robotics, HCI, robotics and social media, AI, robotics and education, image analysis, data processing. He has contributed to the implementation of several algorithms and architectural models to develop various complex systems and has published several journal and conference papers.