On online collaboration and construction of shared knowledge: assessing mediation capability in computer supported argument visualization tools

SCHOLARONE™
Manuscripts

On online collaboration and construction of shared knowledge: assessing mediation capability in computer supported argument visualization tools

Luca Iandoli[a,b], Ivana Quinto[b*], Anna De Liddo[c], Simon Buckingham Shum[c]

a. School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ, USA
b. Department of Industrial Engineering, University of Naples Federico II, Naples, Italy
c. Knowledge Media Institute, The Open University, Milton Keynes, UK

*(corresponding author)

Professor Dr. Luca Iandoli
Department of Industrial Engineering, University of Naples Federico II
80, Piazzale Tecchio, Naples, Italy, 80125
iandoli@unina.it, (0039) 081 7682935, Fax (0039) 081 7682154
School of Systems and Enterprises, Stevens Institute of Technology,
Castle Point on Hudson, 07030 Hoboken, NJ (USA)
liandoli@stevens.edu, (001) 201 216 8104

Dr Ivana Quinto
Department of Industrial Engineering, University of Naples Federico II
80, Piazzale Tecchio, Naples, Italy, 80125
Ivana.quinto@unina.it, (0039) 081 7682932, Fax (0039) 081 7682154

Dr Anna De Liddo
Knowledge Media Institute, The Open University
Walton Hall, Milton Keynes, United Kingdom, MK7 6AA
anna.deliddo@open.ac.uk, (0044) 1908 653591, Fax (0044) 1908 653169

Professor Dr. Simon Buckingham Shum
Knowledge Media Institute, The Open University
Walton Hall, Milton Keynes, United Kingdom, MK7 6AA
simon.buckingham.shum@open.ac.uk, (0044) 1908 653591, Fax (0044) 1908 653169

Abstract:

Collaborative Computer-Supported Argument Visualization (CCSAV) has often been proposed as an alternative over more conventional, mainstream platforms for online discussion (e.g. online forums and wikis). CCSAV tools require users to contribute to the creation of a joint artifact (argument map) instead of contributing to a conversation. In this paper we assess empirically the effects of this fundamental design choice and show that the absence of conversational affordances and socially salient information in representation-centric tools is detrimental to the users' collaboration experience. We report empirical findings from a study in which subjects using different collaborative platforms (a forum, an argumentation platform, and a socially-augmented argumentation tool) were asked to discuss and predict the price of a commodity. By comparing users' experience across several metrics we found evidence that the collaborative performance decreases gradually when we remove conversational interaction and other types of socially salient information. We interpret these findings through theories developed in conversational analysis (common ground theory) and communities of practice and discuss design implications. In particular we propose balancing the trade off between knowledge reification and participation in representation-centric tools with the provision of social feedback and functionalities supporting meaning negotiation.

Keywords: online conversations, online collaboration, knowledge representation, Debate Dashboard, Argument Mapping Tool, common ground.

Introduction

The most popular collaborative web-based tools such as forums, blogs and wikis are very effective in fostering information sharing and collaboration among many distributed participants through online large-scale conversations. However they suffer from several shortcomings when it comes to support online debate, with typical concerns being related to the quality and redundancy of user-generated content (Gurkan, Iandoli, Klein & Zollo, 2010; Kittur, Suh, Pendleton & Chi, 2007; Klein, 2012; Viegas, Wattenberg, Van Ham, Kriss & McKeon, 2004), inability to limit the effects of collective irrationality (Sunstein, 2005; Sunstein, 2006), and to generate knowledge representations that can be easily re-used, e.g. for subsequent expansion, or amendment (Conklin, 2006).

In order to address some of these issues, new online collaborative technologies have been developed for geographically dispersed users to build explicit and reusable shared knowledge representations. In particular, in this paper, we focus on collaborative computer-supported argument visualization (henceforth CCSAV) (Kirschner, Buckingham Shum & Carr, 2003; Okada & Buckingham Shum, 2008). CCSAV platforms aim at helping users in the identification and structuring of complex issues through the collaborative construction of argument maps. An argument map is a visual representation of the structure of an issue in informal logic (Walton, 1996) in which the elements of an argument are displayed in trees or networks, whose nodes and links represent, respectively, the different constituents of an argument (such as issues, claims, premises, evidence) and the argumentative relationships among them (support, attack, warrant, etc.).

Argumentation platforms, by design, foster a representation-centric collaboration process in which the collective effort is channeled toward the creation of a shared map and interaction is constrained within the boundaries of the adopted representation formalism. In argument maps

knowledge organization follows the logic of argumentation and content is positioned across the map based on its argumentative relevance; as a consequence the timeline determined by the turn-taking, reply structure that is typical of ordinary conversations is deliberately suppressed.

In this paper we assess the impact of this design choice on online collaboration; in particular, we report empirical evidence from a controlled experiment in which we compared the collaborative process and outcomes generated by different online groups using three online collaboration technologies: an online forum, an hybrid 'socially augmented' CCSAV tool, a plain CCSAV tool. The online groups were set to the same task that consisted in discussing about the likely price trend of a commodity in the short term. The participants in each condition were asked to rate the perceived quality of the collaboration process and make an individual prediction after the discussion was over.

Our findings show that representation-centric platforms affects negatively users' collaborative experience and provide evidence of a tradeoff between representational advantages and meaningful social participation as measured through a set of collaboration metrics including mutual understanding, perceived effectiveness of collaboration, perceived ease of use, and quality of individual decision.

In the next section, we offer a more detailed comparison between conversation and representation centric tools and discuss their strengths and weaknesses in online debates. We then describe our theoretical framework based on studies in conversational analysis (Common Ground theory, Clark & Brennan, 1991) and Communities of practices (Wenger, 1998). Finally, we present the design and discuss the results of the empirical study.

2. Current web-based technologies for collective deliberation

Nowadays several web-based tools such as forums, wikis, microblogs, etc. offer easy ways to support affordable and fast collection of users-generated content through asynchronous, large-scale, distributed conversations. de Moor and Aakhus (2006) label these platforms *sharing tools* as opposed to *collective mapping tools* and analyze their points of strengths and limitations with respect to their ability to support online debate on potentially controversial issues. In this paper we will refer to them as conversation-centric tools as opposed to representation-centric collaborative platforms.

Conversation-centric tools are web-based tools that allow people to share knowledge among many dispersed users through multiple-voices conversations. The most commonly used web-based tools in this category are wikis, forums, blogs, and more recently micro-blogs and other social media.

Although such tools have been remarkably successful at enabling a global explosion of ideas and knowledge sharing, they face serious shortcomings when used to support debate.

First, conversations often incur in edit "wars" or flamed debates on controversial topics (Klein, 2010; Lee, 2005). Second, content is captured in an unstructured and unsystematic fashion, and as a result it may be scattered, redundant and of arguable quality (Klein, 2010). Filtering for relevant contributions is difficult and time consuming because content is often duplicated and highly fragmented. Unsystematic coverage of a topic makes hard to understand which debate areas are well explored and which would deserve further examination. Questionable quality and irrelevance are responsible for low signal to noise ratio in online conversations. Finally, there is no intrinsic mechanism to support critical thinking and valid argumentation.

Representation-centric tools are designed to support structured interaction and creation of shared knowledge artefacts and should be less prone to the inefficiencies that characterize conversations. Some examples of representation-centric tools include: Mind maps (Buzan, 1979), Concept maps (Novak & Gowin, 1984), Issue Maps (Kuntz & Rittel, 1970) and Argument maps (Wingmore, 1913).

Collaborative Argument Mapping platforms help participants to represent a discussion as a visual maps composed of a set of *issues* to be answered, alternative solutions (*positions* or *ideas*), and pro or con *arguments* associated to the proposed ideas. The debate is summarized into a visual map connecting *Issues, Positions* and *Arguments* through labelled links such as *supports, objects-to, refer-to,* etc.

Diverse rational motivations and expectations can be offered to justify the use of these tools to support online debate.

Argument maps work as a sort of group memory and they support better exploration of the problem space (Okada & Buckingham Shum, 2008), provide representational guidance (Suthers, Vatrapu, Medina, Joseph &Dwyer, 2008) and rational organizations of contents (Gurkan, Iandoli, Klein & Zollo, 2010), support collective sense-making and shared cognition (Conklin, 2003), or dialectics and critical confrontation of opposite points of views (Yearwood & Stranieri, 2006).

CCSAV are supposed to overcome some of the typical dysfunctions of current web-based collaborative tools (i.e. forums), such as redundancy, lack of navigability, scattered content and inability to generate formal knowledge representations (Klein, 2010). Argument-based platforms are also designed to encourage criticisms and comparison of diverse, alternative points of view (Carr, 2003; Yearwood & Stranieri, 2006), to stimulate critical thinking (Twardy, 2004; van Gelder, 2007) and evidence-based reasoning (Carr, 2003).

Notwithstanding the expected or actual advantages, argumentation technology, as well as many other tools aimed at making online interaction more structured and "rational", seems to struggle to reach widespread diffusion both in organizations and online communities. We argue that CCSAV designers have underestimated important social and communication factors that help to make users' experience more engaging and their interaction more socially salient. As evidence to this, there is a lack of studies in CCSAV literature on the assessment of the ability of these tools to mediate interaction effectively, and the majority of works in this area are focused on knowledge representation (e.g. search for adequate argument ontologies, Suthers, 2001; Suthers, 2003) or on the impact of CCSAV on cognitive performances in a variety of tasks and settings including online learning (Nowak & Gowin, 1984; Suthers *et al.,* 2001; Suthers & Hundhausen, 2001; Suthers, Vatrapu, Medina, Joseph & Dwyer.*,* 2008), problem solving (Cho & Johnassen, 2002), and decision making (Karakapilidis & Papadias, 2001; Karakapilidis & Tzagarakis, 2007; Karakapilidis *et al.,* 2009).

Some of the critiques to conversation centric tools and the advantages associated to representation centric tools are founded on an idealized view of collective deliberation and deliberative democracy, which has its roots in Habermas's theory of communicative action (Habermas, 1981): according to such perhaps hyper-rational view, deliberation processes should be cleaned up to get rid of all the noise generated by a process that in the real world often entails confrontation, antagonism, rhetoric rather than logic-based and manipulative argumentation strategies, as well as complex and articulated social dynamics needed to achieve consensus. Not only many of the existing deliberation software platforms do not keep into account this complexity, but they are designed with the deliberate intent to suppress it. In this paper, instead, we argue that at least some of these processes should be seconded; in particular our focus is on the comparison between the collaborative dynamics enabled by

conversational interaction and those that are possibly triggered when the coordination of collective action is achieved through a visual device that is supposed to capture collective knowledge.

Despite the richness and abundance of CCSAV studies, we argue that a critical and not adequately investigated issue in this literature is the use of communication formats that may be too constraining and disruptive of smooth users' interaction.

Indeed, the proficient use of argument mapping tools requires a steep learning curve that can be overcome only through a regular and prolonged amount of practice and training (Twardy, 2004). Argument-based graphical representations can be felt as unnatural and unintuitive compared to other possible ways of coordinating interaction based on the time-based, turn taking structure of conversations (Wikes-Gibb & Clark, 1992). CCSAV tools prioritize formal representation of contents and doing so they disrupt by design the temporal flow and the conversational reply structure that users are familiar with, thus making collaboration more awkward.

While CCSAV are designed to support the construction of rational and reusable knowledge outputs, more empirical analysis is needed to assess whether they have a dysfunctional impact on the collaboration process. In the following section we provide further theoretical justifications for this hypothesis drawing from influential contributions advanced in the literature on communities of practice and conversational analysis.

3. Theoretical framework

3.1 The duality of collaborative knowledge creation process: participation *versus* reification

According to Wenger (1998), knowledge creation in communities of practices requires the interaction of two mutually reinforcing processes: participation and reification. Participation refers to the process of being an active part of some types of ongoing collective action, by interacting with other members and sharing ideas and information, but also goals, resources, means for actions and workload. Reification refers to creating a collective output by embodying collective knowledge and experience into shared artifacts. Through reification concepts and ideas are objectified into something that can be seen, used, operated or manipulated by other participants for various purposes (communication, understanding, learning, negotiation, etc.). Examples of reification are bylaws, organizational charts, contracts, and quality manuals.

Participation and reification are mutually reinforcing and must be properly balanced (Wenger, 1998): if participation prevails, interaction will not lead to the creation of shared artifacts and participants will feel frustrated and perceive the process as sterile; if reification prevails, hard objectification of knowledge will hinder collective action by preventing revisions and re-elaboration, thus ultimately suffocating participation. For instance, a brainstorming session will be perceived as more successful if it will help participants to come up with a plan, a to-do list or any other documentable output containing actionable knowledge representing the group's contribution to address a given problem.

By using technologies such as forums or wikis, people can participate to a conversation, interact, and share information. Interaction could lead to the creation of collaborative knowledge artifacts, such as texts, pieces of software codes, and new designs. The creation of such knowledge objects will affect the subsequent interaction that will consequently become more focused and object-oriented.

Following Wenger (1998), conversational tools, such as online forums and blogs, privilege

participation over knowledge reification. Representation centric tools, including CCSAV

platforms, emphasize instead the creation of knowledge objects and keep participation on the

background, assuming that it will not be significantly affected by map-centric interaction. We

critically re-examine this assumption in the next section, in which we assess CCSAV as a

form of mediated communication.

3.2 The costs of mediating interaction: the Common Ground Theory

Clark and Brennan (1991) claim that in a conversation participants do not only exchange

content, but also different types of feedback, delivered through verbal and nonverbal

communication. Participants rely on such feedback to collect and offer evidence that they

understand each other; once this evidence is available, they can update the amount of mutual

knowledge that they presume to share. The joint process of construction of this mutual

knowledge is called *grounding*; the cumulative stock of mutual knowledge is referred to as

common ground. For example, a speaker perceives simple gestures such as nodding or eye

contact on the part of the listener as signs that the other person is engaged and is getting the

message; on the contrary, raised eyebrows or a direct inquiry through which the listener asks

explicitly for clarification are evidence that the communication problematic.

According to Clark and Brennan (1991), common ground is a necessary condition for a

conversation to proceed and the accumulation of mutual knowledge leads to more

straightforward, efficient, and convenient communication, which facilitates coordination, and

collaboration (Convertino, Ganoe, Schafer, Yost & Carroll, 2005).

When conversations are mediated by a technology, part of the feedback readily available in a

face-to-face setting is either missing or must be provided with some extra effort. For example,

in face-to-face conversations, people can use facial and body expression to manifest

agreement, while in computer-mediated conversations participants should provide the same information by using different techniques, such as emoticons or other types of conventional etiquette.

As predicted by the Common Ground theory (Clark and Brennan, 1991), mediated communication is always less efficient than face-to-face interaction because the latter entails higher grounding costs. One consequence of the theory is that the use of a mediating technology should be encouraged only to the extent that the benefits achievable through mediated communication significantly offset the grounding costs that participants have to bear.

Grounding costs can be estimated on the base of the affordances that are available in a given medium (Clark & Brennan, 1991; Kraut, Fussell, Brennan & Siegel, 2002). While these affordances (table 1) are readily available in face-to-face conversations, in mediated communication they may be partly missing; the higher the number of such affordances, the more efficient the interaction will be because participants will not need to compensate for the missing feedback. In the second column of table 1, we adapt the original definitions provided by Clark *et al*. to web-mediated conversations.

"Insert Table #1"

It is easy to check that in CCSAV-mediated interaction the overall grounding cost is very high because up to eight out of ten of the conversational affordances that help users to develop common ground are missing, namely: co-presence, audibility, visibility, tangibility, mobility, co-temporality, simultaneity and sequentiality. In online forum conversations, more affordances are instead available, which can explain why this medium is highly popular and widespread. For instance, the reply structure that is visible in forums helps to support sequentiality and co-temporaility; also, the engagement of participants into conversational

turns makes visible the conversational acts performed by other participants. Most importantly, by using the reply structure participants can provide immediate feedback or ask for it, for instance through questions. None of these actions are supported by CCSAV tools, which can make their users' experience awkward and impersonal. The lack of conversational and social feedback creates a need for additional external aids, such as facilitators (Gurkan, Iandoli, Klein & Zollo, 2010) or organizational sponsorship and cheerleading (Conklin, 2006) to support adoption and improve participation.

In order to evaluate the amount of grounding cost in CCSAV-mediated collaboration, we developed a new argument mapping tool that is able to offer the traditional argument mapping functionalities along with others aimed at increasing the level of social engagement of the participants: the Debate Dashboard (Iandoli, Quinto, De Liddo & Buckingham Shum, 2014). Debate Dashboard is a socially-augmented platform able to deliver several types of social and conversational feedback through a set of visual add-ons built upon an underlying argumentation tool (Cohere, Buckingham Shum, 2008) to help participants to build a common ground as suggested by Clark and Brennan theory.

The aim of this work is to assess empirically whether and to which extent missing communication affordances due to the use of argumentation formalism may hinder interaction and reduce CCSAV capability to mediate interaction. Following the grounding cost theory, we compare the collaborative performances of three different technologies used for a same problem solving task: i) online forum, ii) socially augmented CCSAV tool (Debate Dashboard) and iii) plain argument mapping tool (Cohere). We hypothesize that such tools can be positioned along a continuum whose extremes are the online forum and the plain argument-mapping tool.

4. Research hypotheses

We ran an experiment with 95 subjects assigned to three groups. The subjects were students in two classes from an undergraduate program in Engineering Management, age 19-22, 57% male. The participants were required to discuss the price trend of given commodities using three different collaboration platforms: online forum, Debate Dashboard, and an argument mapping tool. All subjects were eventually required to make an individual prediction about the price trend of a commodity in the short term after the discussion was over.

The aim of the test was to evaluate whether representation-centric collaborative platforms affect the collaboration process and outcomes. We compared the three technologies with respect to a number of constructs:

a) Mutual understanding (MU), as defined in the common ground theory;

b) Perceived effectiveness of collaboration (PEC), i.e. the degree of users' satisfaction about the collaborative experience;

c) Perceived ease of use (PEU) of the platform.

Finally, we measured the accuracy of individual predictions by comparing subjects' post discussion estimates with the actual market values in the following quarter.

4.1 Mutual Understating and Perceived Effectiveness of Collaboration

Following Wenger (1998), the three platforms (Online Forum, Debate Dashboard and Argument mapping) are characterized by increasing degree of knowledge objectification. Since representation centric tools have less affordances to support common ground building then we hypothesize that constructs such as mutual understanding and perceived effectiveness of collaboration will be higher in the online forum condition and will progressively decrease moving toward representation centric tools in the following order: i) online forum; ii) socially augmented CCSAV (Debate Dashboard); iii) plain CSAV platform.

More specifically, we will test the following hypotheses:

H1: MU will be higher in the online forum situation, more specifically:

H1a: users of the online forum will achieve higher level of MU than users of the socially augmented CCSAV.

H1b: users of the users of the socially augmented CCSAV will achieve higher level of MU than users of the plain CCSAV.

H2: PEC will be higher in the online forum; more specifically:

H2a: users of the online forum will achieve higher level of PEC than users of the socially augmented CCSAV.

H2b: users of the socially augmented CCSAV will achieve higher level of PEC than users of the plain CCSAV.

4.2. Perceived ease of use

Forums are very popular platforms which Internet users are typically familiar with. They support the reply structure of conversations and are easy to use. We expect instead that the combination of novelty and the use of argumentation formalism in mapping tools will have a negative impact on perceived ease of use. However, we expect that the Debate Dashboard could do better than the plain CCSAV platform in terms of perceived ease of use because social translucence (Erickson & Kellogg, 2002) favored by social augmentation should facilitate users' interaction and engagement. So, we hypothesize that:

H3: PEU will be higher in the forum condition; more specifically:

H3a: the users of the online forum will achieve higher level of PEU than users of the socially augmented CCSAV.

H3b: the users of the socially augmented CCSAV will achieve higher level of PEU than the users of the plain CCSAV.

4.3 Accuracy of Prediction

In addition to the impact of the platforms on the collaborative experience, we wanted to assess whether the use of a specific type of tool had also an effect on the ability of its users to make more informed individual guesses after participating to the discussion.

The common ground theory does not provide straightforward implications in terms of the relationship between grounding cost and quality of the outcomes of the interaction. In fact, the link between how users experience the collaboration process and the quality of outcomes of the collaboration appears to be unclear. For instance, there is an abundant literature on group decision-making offering empirical evidence of group failures in face-to-face situations such as meetings and committees (for a review see Sunstein, 2006). In everyday life, individuals frequently experience that conversations may have very different outputs, ranging from great insights, to decision-making failures and pure waste of time, and that enjoying a conversation does not necessarily entail having a productive one.

On the other hand online collaboration removes or at least weakens some of the factors that can be responsible for group decision-making failures: persistence of contents, asynchronous interaction, and possibility of anonymity ensuring limited and "safe" personal confrontation may help participants to be more reflexive and task-oriented, thus weakening some of the social dynamics that can lead a group astray. One may expect that this type of support is even stronger when interaction is mediated by an argument mapping tool, because CCSAV platforms are supposed to help users in locating relevant knowledge faster and allow for the representation of conflicting points of view (Klein, 2010), and foster critical thinking and dialectical inquiry in the community (Grasso, Cawsey & Jones, 2000; Mercier & Sperber, 2009; Stranieri, Zeleznikow & Yearwood, 2001). On the other hand, the presence of higher grounding costs in representation centric tools may make collaboration more awkward, which in turn may reduce participants' engagement and their ability to contribute and access to a shared knowledge base.

Given the problematic nature of the relationship between the quality of the collaborative process, the adopted type of knowledge representation, and the interplay between the way the interaction process is constrained and what a collective is able to produce as a result of this interaction, we do not advance specific hypotheses about the impact of the different platforms on the quality of decision. We simply measured the accuracy of the post debate individual prediction of the price, compared with the actual market data. The measure was elicited through the following question: "what do you think will be the price trend of the commodity three months from now?".

5. Method

5.1 Study design

Since the objective of this study was to compare the users' performance of three different platforms in a same task, we adopted a between-subjects 3 x 1 experimental design. Following Van de Braak et al. (2006), in order to avoid undesirable influence of pre-existing differences between subjects we performed a random assignment of subjects to the three conditions and performed a statistical test to check that the groups were not significantly different with respect to demographics (gender, age) and academic performance; the results showed that the differences between the three groups were not significant [$F(2, 162)=1.210$, $p>.05$].

5.2 Participants and Task

165 subjects were initially recruited for the empirical study. All subjects were undergraduate students belonging to two different classes of a same course in Economics (Micro and macro economics), age 19-22, 57% male; participation was compensated with extra credits. Eventually only 95 students completed all the activities (the majority of those who dropped did so before the start of experiment, because of technical difficulties in the installation of a

Virtual Machine needed to monitor the navigation sessions or for personal reasons). Finally, the three groups were so composed: 34 in group A (online forum), 25 participants in group B (socially augmented CCSAV - *Debate Dashboard*), and 36 subjects in group C (plain argument mapping tool - *Cohere*). A fourth group performed the same task in a face-to-face discussion but this further data was not used in the analysis.

Due to the higher number of defections, we rechecked that the three groups still were not significantly different in terms of demographic and academic performance, which turned out to be the case [$F(2, 91) = 1.108$, $p > .05$].

The subjects were asked to discuss about what would be the future price of a given commodity in the short term (three months from the end of the discussion) and make an individual forecast of its price after the discussion was closed.

We used the following criteria for the determination of the type of task:

o *Realism and information richness*: participants were meant to face a real-world decision task rather than an abstract choice problem, as those used in lab experiments. Moreover, the task had to be controversial and complex enough to produce an adequate level of discussion and information exchange around the topic.

o *Acceptable task difficulty*: task difficulty had to be compatible with the skills and background of the participants.

o *Measurability of the outcome*: we looked for a task with a unique correct solution against which to compare the predictions made by each individual participant after the debate.

As cases for discussion we selected two commodities (Crude Oil and Gold). Predicting the price of these goods is *realistic* and relevant for Economists and Business operators. Crude oil

and gold prices are affected by many variables; their forecast is notoriously a difficult task and requires an extensive analysis that could benefit by the collective exploration of a large and uncertain decision space. While the problem has a unique correct answer that can be ascertained ex-post, the prediction of prices of sensitive commodities is hard due to the many factors influencing market equilibrium and the complex relationships among them.

As far as the alignment between task difficulty and participants' skills, while the students were obviously not domain experts, their attendance of the Economics course provided them with some background on market equilibrium, good enough to produce at least an educated guess. Besides, because of their impact on the Economy, it is not difficult to find information about crude oil and gold prices in the news, experts' analysis, and even through direct experience in everyday life. Additionally, information and readings were provided during the course for the specific case.

Finally, the task satisfies the *measurability* criterion, because crude oil and gold prices historic trends are available for later comparison with the subjects' forecasts.

Students were divided in three groups to which crude oil and gold were respectively assigned; in particular, group A and B discussed on the future price of gold, while the group C was required to forecast the price of crude oil. All the subjects had to perform the same task: predicting the market price of a different commodity. The reason why we assigned two different commodities to the three groups was that two of the three groups, namely group B and C, were in a same class and we wanted to reduce information spill over and collaboration among students belonging to diverse groups.

In principle this choice may affect the students' performances and their outputs, but the two tasks can be considered similar enough in terms of nature and level of difficulty, also

considering that the students were not experts in either cases and were required to apply the same general model of market equilibrium presented in the class. Additionally, we provided students in each condition with reading materials and information that were similar in terms of sources, quantity and depth of analysis. Our empirical data show that the two tasks attracted similar level of participation, as measured by the numbers of posts (351, 269, and 334 respectively for group A/Online forum, group B/Augmented CCSAV and group C/Plain CCSAV).

In the follow-up questionnaire we asked students to highlight possible problems and difficulties both with the platforms and with the task, as well as to provide some suggestions and feedback, but we did not find differences in terms of perceived task difficulty related to the topic. Finally, we double-checked through historical data (http://www.gold.org/investment/interactive-gold-price-chart; http://www.oecd.org; http://www.nasdaq.com/markets/crude-oil.aspx) that in the considered period none of the two markets had showed any turbulence due to singular events such as external shocks or market hype that could have made the prediction harder in one case.

We set up a two-week online asynchronous discussion in which the members of the 3 online groups were asked to discuss about the price trend of the assigned commodity in the short term using 3 different collaborative tools. The participants received information and reading materials two weeks before the start of the experiment and were instructed to use this time to familiarize with the problem and the domain, search for additional information, and prepare for the discussion.

5.3 Online Collaborative platform

Three different platforms were made available to the participants in the three online groups:

- An *online collaborative forum.* This was built ad hoc using a popular and free social media platform (we used a Ning forum http://www.ning.com/)

- A light version of *Cohere* (Buckingham Shum, 2008) argumentation tool

- The *Debate Dashboard* (Iandoli, Quinto, De Liddo & Buckingham Shum, 2014), for the socially augmented CCSAV platform;

"Insert Figure #1"

Cohere is a web-based asynchronous argument mapping tool whose purpose is to support on-line argumentative debate (http://cohere.open.ac.uk/) (Buckingham Shum, 2008; De Liddo & Buckingham Shum, 2010). The platform is implemented on Linux, Apache HTTP server, My SQL database and PHP. Cohere is currently an Open University's (OU) legacy project (started in 2006, the freely hosted application was discontinued in Nov 2014 and in now available through OU's intranet only) and an open source software. With Cohere users can create posts to express their thoughts, and they can associate icons to them, in order to explain the rhetorical role of each post in the wider discussions (Fig. 1). Moreover, with Cohere users can explicitly connect posts and express the argumentative relationship between posts with a positive, neutral and negative semantic link. By structuring and representing online discourse as semantic network of posts, Cohere allows users to break the discourse down into an externalised and sharable artefact, the argument map, which enable users to better analyse and visualize the pro and con of each prediction and to better make sense of discourse contents.

The Debate Dashboard is a socially augmented version of Cohere (for a detailed description see Iandoli, Quinto, De Liddo & Buckingham Shum, 2014) that uses Cohere's argument mapping functionalities and integrates them with a number of additional widgets to improve visibility of users' identity, online presence and contributions to the debate. For instance it is possible to visualize personal information including users profile, active presence on the site,

and contributions to the discussion. A social graph can be launched to show the connections between members, as determined by their interaction over the platform. Social tagging makes possible for users to emphasize keywords, which are then displayed through a navigable tag cloud that reflects the current popularity of keywords and topics. These and other visualizations have been designed to reduce the loss of social and conversational feedback and, consequently, to prevent the increase of grounding costs. Figure 2 shows a Debate Dashboard page, where users can access easily the visual feedback through different tabs, such as the social network visualization, find out which are the most used tags, who are the other members (People tab) and what they do in the discussion (Stat tab).

In previous work (Iandoli, Quinto, De Liddo & Buckingham Shum, 2014) we describe more in detail how we identified the added functionalities through the analysis of the communication affordances identified in the Common Ground Theory (table 1). This description is out of the scope of this work and we refer the readers to our previous publication for more details.

"Insert Figure #2"

5.4 Procedures and data collection

The test was carried out in three steps: i) preparatory work, ii) a two-week discussion/preparation, iii) follow-up questionnaire.

Before the two weeks, all the students using the online argumentation tools had four 2-hours seminars about: a) Collective intelligence and online collaboration applications, b) Argumentation theory, with a focus on IBIS and argument-based tools, c) the Gold and Crude Oil Markets, d) an instructional demo of the online collaborative argument-based platforms. The students in the forum participated only to a) and c) and received instructions to register and access to the forum. All the students had access to a few reading materials and web links from

specialized sources, including financial and economic newspapers and magazines, blogs, governmental organizations websites, etc. The users of the argument mapping tools went through a warm up phase for one week during which they had the opportunity to familiarize with the platforms discussing on a different topic.

A framing question and alternative answers were provided to the participants in each condition: what do you think will be the trend of gold/crude oil price in three months from now? The possible answers were: i) The price will tend to increase (+10% or higher), ii) The price will tend to decrease (-10% or lower), and iii) The price will be stable (+/-10%).

The experiment was run in an asynchronous fashion for two reasons. First, we wanted our subjects to have enough time to explore and reflect on the issue; second, CCSAVs are typically used to support asynchronous discussions.

5.5 Measures

The performance variables are measured through a 24 questions survey aimed at evaluating three constructs at the individual level: Mutual understanding, Perceived quality of collaboration, and Perceived ease of use. As for the quality of individual decision we simply polled the subjects for an individual forecast and compared those forecasts with the actual data three months later. In particular:

- *Perceived Effectiveness of Collaboration* was measured with a set of 8 questions identified by reviewing the literature on mediated collaboration (Daily-Jones, Monk & Watts, 1998; Sellen, 1995; Vandergriff, 2006). The 8 questions aimed at measuring the efficacy and quality of the collaboration process, the users' satisfaction about the discussion and the interaction, as well as the users' perception about the amount of knowledge sharing during the experiment.

- *Mutual Understanding* was measured through 9 items defined on the basis of a literature review on the Grounding cost theory (Clark & Brennan, 1991) and on studies about common ground building in mediated conversations (Monk & Watts, 2000; Convertino *et al.*, 2007; 2008; 2009). The questionnaire's items assess the level of mutual understanding, the efficacy of social and conversational feedback in supporting mutual understanding, the redundancy and the irrelevance of created posts as proxy for the lack of mutual understanding (Whittaker, Terveen, Hill & Chemy, 1998).

- *Perceived ease of use* was assessed through 6 items based on Davis' Technologies Acceptance Model (1989), and its subsequent modifications and extensions (Daily-Jones, Monk & Watts, 1998; Vassileva & Sun, 2007; Venkatesh, 2000; Venkatesh & Davis, 2000).

We assessed the constructs in terms of reliability and validity. As far as reliability is concerned, we obtained acceptable values of the Cronbach's alpha for each construct (Table 2).

"Insert Table #2"

Table 3 shows the results of the discriminant validity. As the square root of the average variance extracted is much larger than the correlations of the construct with the all of the other constructs, it can be considered an acceptable criterion to pass the discriminant validity test (Fornell & Larcker, 1981).

"Insert Table #3"

As far as the variable Accuracy of Prediction is concerned, we carried out a poll at the end of online discussion in which all subjects were asked to answer to the following question: "what do you think will be the price trend of the commodity three months from now?", selecting one

of three mutually exclusive options ("the price will be 10% higher than today", "the price will 10% lower than today", "the price will stay around the present value"). We then compared the individual forecasts with the actual market data to assess the accuracy of prediction.

6. Results

In order to compare the collaboration platforms, we performed statistical tests to verify whether the three groups differed significantly with respect to the observed variables.

All the variables passed the normality test (Shapiro –Wilks, with $p>0.05$), thus we used one-way Anova analysis to verify the following set of hypotheses (H_0 being the null hypothesis, and *Forum*, *DD* and *AMT* standing, respectively for online Forum, Debate Dashboard, plain Argument Mapping Tool):

Mutual Understanding: $\mathbf{H_0}$: $\mu_{MU Forum}=\mu_{MU DebateDashboard}=\mu_{MU AMT}$
                            $\mathbf{H_1}$: $\mu_i \neq \mu_j$ *for some i and j*

Perceived Effectiveness of Collaboration: $\mathbf{H_0}$: $\mu_{PEC Forum}=\mu_{PEC DebateDashboard}=\mu_{PEC AMT}$
                            $\mathbf{H_2}$: $\mu_i \neq \mu_j$ *for some i and j*

Perceived Ease of Use: $\mathbf{H_0}$: $\mu_{PEU Forum}=\mu_{PEU DebateDashboard}=\mu_{PEU AMT}$
                            $\mathbf{H_3}$: $\mu_i \neq \mu_j$ *for some i and j*

Before proceeding to the ANOVA test, we verified the violation of nonindependence statistical assumption. According to several studies on small groups' research (Kenny and Judd, 1986; Kenny *et al.,* 1998; Kenny *et al.,* 2002), if the individual is used as the unit of analysis, the assumption of independent units is likely to be violated when the task requires intra-group interaction because the subjects may influence each other. A group effect occurs if the score of individuals within a group are more similar to one another than the scores of individuals in different groups; this leads to high likelihood of *nonindependence* of observations. In order to assess and estimate the degree of nonindependence, we computed

the *intraclass correlation* for each dependent variable (Kenny *et al.,* 2002) and tested for

statistical significance (table 4).

<p style="text-align:center">*"Insert Table #4"*</p>

As in this case the F statistics are significant, it is possible to claim that there is

nonindependence. In this case it is not possible to perform directly the ANOVA test as such

because the p-values may be inflated. We therefore used Kenny *et al.* (1998) three-steps

procedures in order to quantify the effect of nonindependence on the ANOVA p-values and

perform ANOVA with adjusted p-values that take into account this inflation: i) a critical value

for the F test with individual as units of analysis is determined for the available degrees of

freedom; ii) such critical value is divided by a bias factor to produce an adjusted F; iii) the

degrees of freedom for the adjusted F test are reduced on the basis of the type of independent

variables and the size of the intraclass correlation. The adjusted F can be used to compute the

effective alpha values for the ANOVA test. Table 65 reports the results of this test, which

clearly show that, due to the nonindependence, the *p-values* are grossly inflated.

<p style="text-align:center">*"Insert Table #5"*</p>

Tables 6 and 7 show the results of the one-way ANOVA test that we performed to check

whether the differences among the 3 groups in terms of MU, PEC and Perceived ease of use

are still statistically significant under the adjusted p-values.

<p style="text-align:center">*"Insert Table #6"*</p>

Therefore, after discounting for the inflation determined by nonindependence, there is still a

statistically significant difference between groups as whole as determined by one-way

ANOVA with respect to all three variables.

Table 8 reports the results of the post-hoc tests, namely the mean differences among all groups

for each construct using the adjusted p-values reported in table 5.

"Insert Table #7"

The Tukey post-hoc test results (Bowker & Lieberman, 1959) show that there are significant

differences in terms of mutual understanding, perceived quality of collaboration and

perceived ease of use between the group that used the online forum and the group that used

the plain AMT (respectively $p=.000$, $p=.000$, $p=.0029$). However, there are no significant

differences between the groups that used the online forum and those that used the Debate

Dashboard as well as between students that used the Debate Dashboard and those that used

the plain AMT for each of the three independent variables.

Though not all the differences are statistically significant, the Tukey post-hoc test confirmed

that the group that used the plain argument mapping tool (AMT) obtained the lowest score

with respect to each dependent variable (MU: 27.92 ±4.33, $p=.000$; PEC: 38.78 ±4.40,

$p=.000$; PEU: 28.21 ±5.06, p=.0029), compared to the other users of the Debate Dashboard,

who got an intermediate score (MU: 31.12 ±4.68; PEC: 42.08 ±4.02; PEU: 31.96 ±3.61), and

those of the forum, who obtained the highest score (MU: 34.15 ±3.69; PEC: 45.09 ± 4.45;

PEU: 33.38 ±7.19)].

According to these results the three technologies can be positioned along a continuum ranging

from more conversational-centric tools (online forums) to representation-centric tool (pure

CCSAV), along which collaborative metrics such as mutual understanding and perceived

effectiveness of collaboration decrease (figures 3, 4 and 5).

"Insert Figure #3"

"Insert Figure #4"

"Insert Figure #5"

We finally tested for the existence of a significant difference among the groups in terms of accuracy of individual predictions. Since this is a binary variable we used a non-parametric test (Friedman test). A statistically significant difference between the groups as a whole ($\chi^2(2)$ = 27.636, p = .000), was found. Figure 6 shows that the mean value of the accuracy of decision follows the same trend observed with the previous constructs, decreasing as one shifts from the forum toward collaborative mapping tools.

"Insert Figure #6"

7. Discussion

The aim of this work was to assess whether and to which extent representation-centric interactions influence users' collaborative experience and performance. More specifically we wanted to evaluate CCSAV platforms in terms of their ability to mediate interaction from a user-centric perspective. Our study offers a novel contribution to CCSAV literature in which the focus is on the impact of argument-based knowledge representation and visualization on group coordination.

Overall, our findings suggest that more formal and representation-centric tools may be less supportive of effective interaction. In particular we show that the losses of mutual understanding, perceived effectiveness of collaboration, perceived ease of use as well as of accuracy of individual post-debate predictions increase progressively the more the platform is representation-centric. As the mean differences between the Debate Dashboard and the plain argumentation platform are not statistically significant after correcting for non independence,

it is unclear if adding social translucence to argument map-centric conversations can help to offset grounding costs.

One interpretation of these findings is that argument mapping tools introduce some serious disruption in the ways people communicate: the lack of transparency and social visibility over other participants' social presence, the difficulty of creating a common ground without asking for and providing feedbacks through the reply structure that is available in conversations, and the difficulty to learn and use a new formalism make users experience socially awkward and less rewarding than ordinary conversational interaction.

A steep learning curve and increased grounding costs may both represent severe barriers for users, however only the former has received adequate and explicit attention in the literature, with several studies reporting about successful accounts of CCSAV applications thanks to the adoption of some mechanism to offset learning costs, such as extrinsic incentives and use of community moderators (Gurkan, Iandoli, Klein & Zollo, 2010), extensive training in the use of argumentation formalism (Twardy, 2004; Van Gelder, 2007), or strong internal cheerleading and organizational sponsorship (Conklin, 2006).

Unfortunately our results are not conclusive to ascertain if the provision of additional socially salient meta-information through hybrid solutions as the Debate Dashboard could be the way to better manage the trade-off between participation and knowledge reification in representation centric platforms. However we think that this study offers a useful and novel theoretical perspective based on the analysis of the gap between conversation and object-centric interaction that can provide some insights and guidelines for the design of better collaborative platforms.

While the provision of social information can help users to be more aware of the social landscape in which they interact, what is currently missing in the Debate Dashboard is some mechanism to support the iterative and dynamic process through which participants can refine

and negotiate meaning to update their mutual knowledge, functioning as a substitute of the conversational reply structure. This limitation may contribute to explain why the Debate Dashboard was not powerful enough to outperform the online forum and the plain CCSAV in terms collaborative performances as we hoped.

Our study has relevant implications for designers of collaborative platforms. A better trade off between participation and reification should be a desirable objective for developers and could be a useful design principle to deliver a satisfying compromise between the need to obtain reusable and more formal collective knowledge representations and the necessity to make interaction smooth and meaningful for the users. The reduction of grounding costs can improve users' collaborative experience and help them to reap the benefits offered by CCSAV platforms in terms of knowledge organization, localization, visualization and re-use.

Our study does not confirm the expectations of the superiority of CCSAV platforms over forums in improving individual forecasts. This result could be in part due to the fact that the students were not domain experts, nor they were trained to use the argumentation formalism. There is evidence from well-known studies in informal reasoning that human beings exhibit limited argumentation skills (Kuhn,1991) in everyday reasoning; additional evidence from CCSAV literature shows that when exposed to proper and extensive training users show significant improvements in their argumentation mapping skills, and experience a positive impact on some cognitive abilities such as critical thinking (Twardy, 2004). Due to the lack of adequate training the participants might not have been in the position to fully reap the benefits of argument-based knowledge representation, and the results in terms of decision outcomes could have been different with a more qualified population of users.

8. Conclusions

8.1 Contribution

In this paper we have presented an empirical study aiming at comparing different types of online collaborative platforms in terms of their ability to mediate collaboration effectively in a collective forecasting task. We have found evidence that the transition from conversational platforms (forums) to representation-centric tools (CCSAV) is characterized by a deterioration of the users' collaborative experience in terms of mutual understanding, perception of the quality of the collaboration process, and of perceived ease of use. We have found also evidence about the impact of the different types of collaborative technology on users' ability to make more accurate forecasts after participating to the online debate, with the forum surprisingly outperforming argument-based tools in this case.

We also found that the users of the Hybrid CCSAV reported more limited deterioration of the collaborative performance compared to those using the plain argumentation platform, but the differences in the pairwise comparison among the tools were not significant after we discounted the nonindependence effect from the ANOVA results.

While we are aware that the statistical power of our analysis is limited, we think that the investigated research question is very relevant, it is addressed in a novel way, and that we found some insights backed by reasonable, although non-conclusive evidence; finally, we propose a rigorous experimental design that could be replicated by other scholars with more groups and resources.

8.2 Limitations and future work

Our study suffers from the common limitations due to the involvement of academic students as subjects for our experiment. While the students had some background to address the problem and were in the position to make educated guesses, they were certainly not experts and it is not straightforward to generalize these results to more qualified users. While on one side it is obvious to expect that deeper domain knowledge and expertise should produce better

forecasts, the issue relevant to this paper is whether or not the availability of an external knowledge representation can help to improve collaboration and to some extent make better decisions, and if the effectiveness of this type of support depend on subjects' level of expertise and proficiency in the specific task. For instance, representational guidance (Suthers, 2003) helps novices more than experts because the latter rely on internalized representations that they have built over years of field experience. This is an interesting research question that could be pursued in future studies.

Another limitation of our experiment was the use of only one group for each condition. Kenny *et al.* (1998) show that if there is nonindependence then groups and not people should be used as unit of analysis. Consequently, we undertook several actions to maximize the statistical significance of our findings given the limited number of groups. First, we created the groups through block randomization to at least reduce some of the effects of group composition. Second, we tested for nonindependence and we used a conservative approach for the ANOVA test by assuming the adjusted p-values as reference for the test, as suggested by Kenny et al. (1998). Third, all the metrics we considered in our tests are post-experiment individual judgments provided by subjects through questionnaires that were filled between one and two weeks after the experiment; while this is not enough to entirely discount the effects of intra-group influence, it is reasonable to assume that introducing some time lag between the experience and its assessment by individuals should at least attenuate the effect of nonindependence.

However, attempting to adjust for nonindependence is not an ideal solution. Instead of applying statistical correction after the fact, it would have been better either to recruit additional groups and consequently more participants, or to divide the recruited participants into smaller groups

Additional concerns related to the generalization of our findings arise from the circumstance that our results have been obtained with reference to a specific type of task (forecasting the price trend of a commodity). While this task is relevant and realistic, we cannot offer any evidence that the same outcomes would be obtained with other problems or with tasks of different nature (e.g. creative tasks).

It would be interesting to replicate our study with different types of mapping platforms (concept mapping, causal mapping, etc.). Based on our findings and on the common ground theory we expect to observe the same types of problems also with other representation-centric tools. However, other types of representations, such as causal maps, could be closer to how individuals actually organize knowledge, as demonstrated by studies in cognitive science and discourse analysis, which show that causal representations are more effortlessly understood (Trabasso & Van De Broek, 1985), can be easily manipulated by individuals to index memory and explain reality (Shanck, 1995), and exploited to make decisions in practical situations (Pennington & Hastie, 1990). The investigation of the effects of different types of knowledge representations in online, distributed collaboration represents an unexplored and promising area of study that deserves more attention in our research community.

Though we cannot offer conclusive evidence, in our experiment the social augmentation of an existing CCSAV through the inclusion of functionalities aimed at providing social and conversational feedback on the top of a representation-centric tool did not help to offset the grounding costs determined by the argumentation platform.

One of the rationales for introducing social augmentation was to better balance participation and knowledge reification, as theorized in Wenger's theory. Since we developed a minimum viable prototype for the empirical test, it is possible that better and more refined implementation of the social widgets could have helped to create a platform that is better able to support meaningful social participation. Unfortunately, we cannot offer data based on users'

feedback or actual behavior to evaluate the quality of the implementation of the visual

widgets, nor we have compared alternative design solutions. Future users' evaluation studies

are needed in order to inform and support better user interface.

Besides, in order to contain development costs, we implemented only a limited subset of the

conversational affordances suggested by Clark and Brennan and outlined in table 1.

A second and maybe more important reason is that the type of social feedbacks that were

delivered through the Debate Dashboard support more social awareness rather than

participation; with social awareness we mean that participants become more knowledgeable

of the social landscape in their community by receiving additional information about the

social visibility of their own and others' actions in the virtual environment. We speculate that

increased awareness, while contributing to social translucence (Erickson & Kellogg, 2002),

are not directly supportive of several other aspects that make social participation more intense

and meaningful, such as engagement, social recognition, sense of membership, and social

exchange. Further research is needed to better characterize participation in virtual

communities and identify design solutions that are able to support these needs effectively.

Finally, the improved performance of the forum in terms of collaborative support as measured

through our metrics is probably an evidence of the fact that reification and participation are

not appreciated as equally important by users: when confronted with this trade-off apparently

users enjoy more a platform supporting higher engagement and limited knowledge

objectification capability than the opposite. If this preference is based on deep cognitive

routines, the designers of online collaborative platforms should provide more effort to second

these attitudes and carefully consider the trade off of getting rid of some of the key

components of conversation-centric interaction.

Overall, this study makes a point about the need to rigorously assess the impact of

representation-centric collaborative platforms and take into appropriate consideration the

social loss that is implied by the disruption of more natural interaction styles based on conversations; further research is needed to validate this finding through larger scale studies and identify the design solutions that can help to mitigate or eliminate such a loss while preserving the ability of representation centric tools to generate semi-formal and reusable collective knowledge representations.

Reference

Bowker, A.H. & Lieberman, G.J. (1959). Engineering Statistics, Englewood Cliffs, N. J.: Prentice-Hall, Inc., pp. 286-364.

Buckingham Shum, S. (2003). The roots of computer supported argument visualization. In P.A. Kirschner, S. Buckingham Shum, & C.S. Carr (Eds.) Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making, (pp. 3-24), London: Springer-Verlag.

Buckingham Shum, S. (2008). Cohere: Towards Web 2.0 Argumentation. Proceedings COMMA'08: 2nd International Conference on Computational Models of Argument, (pp. 97–108), 28-30 May 2008, Toulouse, France.

Buzan, T. (1979) Use both sides of your brain. New York: E.P. Dutton.

Carr, C.S. (2003). Using computer supported argument visualization to teach legal argumentation. In P.A. Kirschner, S. Buckingham Shum, & C.S., Carr, (Eds.) *Visualizing Argumentation – Software Tools for Collaboration and Educational Sense-making*, (pp. 75-96), London: Springer-Verlag.

Cho, K.L., & Jonassen, D.H. (2002). The effects of argumentation scaffolds on argumentation and problem solving. *Educational Technology Research and Development*, 50(3), 5-22.

Clark, H.H., & Brennan, S.E. (1991). Grounding in communication. In L.B.L. Resnick, J.M. Teasley & D. Stephanie (Eds.), *Perspectives on socially shared cognition,* (pp. 127-149), Washington (DC): American Psycho Kieslerlogical Association.

Conklin, J. (2003). Dialog Mapping: Reflections on an Industrial Strength Case Study. In P.A. Kirschner, S. Buckingham Shum, & C.S., Carr (Eds.), *Visualizing Argumentation – Software Tools for Collaboration and Educational Sense-making* (pp. 117-131), London:, Springer-Verlag.

Conklin, J. (2006). *Dialogue Mapping: Building Shared Understanding of Wicked Problems*. Chichester (UK): Wiley.

Convertino, G., Ganoe, C.H., Schafer, W.A., Yost, B., & Carroll, J.M. (2005). A Multiple View Approach to Support Common Ground in Distributed and Synchronous Geo-Collaboration. In *Group'07 Proceedings of third International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV)* (pp. 121–132), Zurich, Swizterland.

Convertino, G, Hong, L., Nelson, L., Pirolli, P., & Chi, E.H. (2009). Activity Awareness and Social Sensemaking 2.0: Design of a Task Force Workspace. In *Lecture Notes in Computer Science, Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience,* pp. 128-137.

Convertino, G., Mentis, H.M., Rosson, M.B., Carroll, J.M., Slavkovic, A., & Ganoe, C.H. (2008). Articulating common ground in cooperative work: content and process. In *CHI'08, Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems,* ACM New York, NY, USA.

Convertino, G., Mentis, H.M., Ting, A.Y.W., Rosson, M.B., & Carroll, J.M (2007). How does common ground increase?. In *Proceedings of the 2007International ACM conference on Supporting group work*, pp. 225-228, New York, NY, USA

Daly-Jones, O., Monk, A., & Watts, L. (1998). Some advantages of video conferencing over high quality audio conferencing: Fluency and awareness of attentional focus. *International Journal of Human-Computer Studies*, 45, 21-58.

Davis, F.D. (1989). Perceived usefulness, perceived ease of use and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-339.

De Liddo, A., & Buckingham Shum, S. (2010). Capturing and Representing Deliberation in Participatory Planning Practices. *International Conference on Online Deliberation* (OD2010).

de Moor, A., & Aakhus, M. (2006). Argumentation support: from technologies to tools. *Communications of the ACM*, 49(3), 93-98.

Erickson, T., & Kellogg, W.A. (2002). Knowledge Communities: Online Environments for Supporting Knowledge management and its Social Context. In Ackerman M.A., Pipek, V., & Wolf, W. (Eds.) *Beyond knowledge management: Sharing expertise*, pp. 229-236, MIT Press, Cambridge, MA.

Fornell, C., & Larcker, D.F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, 39-50.

Grasso, F., Cawsey, A., & Jones, R. (2000). Dialectical Argumentation to Solve Conflicts in Advice Giving: A Case Study in the Promotion of Healthy Nutrition". *International Journal of Human-Computer Studies.* 53 (6), 1077-1115.

Gürkan, A., Iandoli, L., Klein, M., & Zollo, G. (2010). Mediating debate through on-line large-scale argumentation: Evidence from the field. *Information Sciences*, 180, 3686-3702.

Habermas, J. (1981). The Theory of Communicative Action: Reason and the Rationalization of Society, Volume 1 (Thomas McCarthy, Trans.). Boston, MA: Beacon Press.

Iandoli, L., Klein, M., & Zollo, G. (2007). Can we exploit collective intelligence for collaborative deliberation? the case of the climate change collaboratorium. MIT Sloan Research Paper No. 4675-08.

Iandoli, L., Quinto, I., De Liddo, A., & Shum, S. B. (2012). A debate dashboard to enhance online knowledge sharing. *VINE*, *42*(1), 67-93.

Iandoli, L., Quinto, I., De Liddo, A., & Buckingham Shum, S. (2013). Socially-Augmented Argumentation Tools: Rationale, Design and Evaluation of a Debate Dashboard. *International Journal of Human-Computer Studies,* 72(3),298–319.

Karacapilidis, N., & Papadias, D. (2001). Computer supported argumentation and collaborative decision making: the HERMES system. *Information systems*, 26(4), 259-277.

Karacapilidis, N., & Tzagarakis, M. (2007). Web-based collaboration and decision making support: A multi-disciplinary approach. *International Journal Web-Based Learning and Teaching Technologies*, 2(4), 12-23.

Karacapilidis, N., Tzagarakis, M., Karousos, N., Gkotsis, G., Kallistros, V., Christodoulou, S., Mettouris, C., & Nousia, D. (2009). Tackling cognitively-complex collaboration with CoPe_it!. *International Journal of Web-Based Learning and Teaching Technologies*, 4(3), 22-38.

Kenny, D.A., Mannetti, L., Pierro, A., & Livi, S. (2002). The Statistical Analysis of Data From Small Groups. *Journal of Personality and Social Psychology,* 83(1), 126-137.

Kenny, D.A, Kashy, D., & Bolger, N. (1998). Data Analysis in Social Psychology. In: Gilbert, D., Fiske, S. and Lindzey, G. (Eds.), *Handbook of social psychology* (4th ed.), 233-265, New York: McGraw-Hill.

Kirschner, P.A., Buckingham Shum, S., & Carr, C.S. (2003). Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making, Springer-Verlag.

Kittur, A., Suh, B., Pendleton, B. A., & Chi, E. H. (2007). He says, she says: conflict and coordination in Wikipedia. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, pp. 453-462.

Klein, M. (2010). Using metrics to enable large-scale deliberation. In Collective Intelligence In Organizations: A Workshop of the ACM Group 2010 Conference (pp. 103-233).

Klein, M. (2012). Enabling large-scale deliberation using attention-mediation metrics. Computer Supported Cooperative Work (CSCW), 21(4-5), 449-473.

Kraut, R.E., Fussell S.R., Brennan S.E., & Siegel, J. (2002). Understanding Effects of Proximity on Collaboration: Implications for Technology to Support Remote Collaborative Work. In Hinds P., and Kiesler S. (Eds.) *Distributed Work,* MIT press, 137-162.

Kuhn, D. (1991). *The skills of argument.* New York: Cambridge University Press.

Kuntz, W., & Rittel, H.W. (1970). Issues as elements of information systems. Working Paper N° 131, Center for Planning and Development Research, Berkeley, USA.

Lee, H. (2005). Behavioral strategies for dealing with flaming in an online forum. The Sociological Quarterly, 46(2), 385-403.

Mercier H. & Sperber D. (2009) Intuitive and reflective inferences. In: Evans J, Frankish K (Eds) In two minds: Dual Processes and Beyond. Oxford University Press, New York.

Monk, A., & Watts, L. (2000). Peripheral Participation in Video-Mediated Communication. International Journal of Human-Computer Studies, 52 (5), 933-958.

Novak, J.D., & Gowin, D.B. (1984). Learning How to Learn. Cambridge University Press, UK.

Okada, A., & Buckingham Shum, S. (2008). Evidence-based Dialogue Maps as a research tool to investigate the quality of school pupils' scientific argumentation. *International Journal of Research and Method in Education*, 31(3), 291-315.

Pennington, N., & Hastie, R. (1993). Reasoning in explanation-based decision making. *Cognition*, *49*(1), 123-163.

Schank, R.C. (1995). The structure of episodes in memory. *Computation and intelligence*, 235-259.

Sellen, A.J. (1995). Remote conversations: The effects of mediating talk with technology. *Human-computer interaction*, *10*(4), 401-444.

Stranieri, A., Zeleznikow, J., & Yearwood, J. (2001). Argumentation structures that integrate dialectical and non-dialectical reasoning. *The Knowledge Engineering Review*, *16*(4), 331-348.

Sunstein, C. R. (2005). Group judgments: Statistical means, deliberation, and information markets. *New York University Law Review*, *80*, 962.

Sunstein, C.R. (2006). *Infotopia*. New York: Oxford University Press.

Suthers, D.D. (2001). Towards a systematic study of representational guidance for collaborating discourse. *Journal of Universal Computer*, 7(3), 254-277.

Suthers, D.D. (2003). Representational guidance for collaborative inquiry. In Andriessen, J., Baker, M. & Suthers, D., (Eds.) *Arguing to learn: Confronting cognitions in computer-supported collaborative learning environments,* (pp. 27-46) Springer, Netherlands.

Suthers, D.D., & Hundhausen, C. (2001). Learning by constructing collaborative representations: An empirical comparison of three alternatives. In P. Dillenbourg, A. Eurelings, K. & Hakkarainen (Eds.) *European Perspectives on Computer-Supported Collaborative Learning, Proceedings of the 1st European Conference on Computer-Supported Collaborative Learning*, (pp. 577-592), Indiana University, Bloomington IN.

Suthers, D.D., Connelly, J., Lesgold, A., Paolucci, M., Toth, E., Toth, J. *et al.,* (2001). Representational and Advisory Guidance for Students Learning Scientific Inquiry. In K.D. Forbus, & P.J. Feltovich (Eds.) *Smart Machines in Education: The Coming Revolution in Educational Technology* (pp. 7-35), AAAI/MIT press, Cambridge, Massachusetts.

Suthers, D.D., Vatrapu, R., Medina, R., Joseph, S. & Dwyer, N. (2008). Beyond threaded discussion: Representational guidance in asynchronous collaborative learning environments. *Computers&Education*, 1103-1127.

Trabasso, T., & Van Den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of memory and language*, *24*(5), 612-630.

Twardy, C. (2004). Argument Maps Improve Critical Thinking. *Teaching philosophy*, 2, 95-116.

van den Braak, S.W., van Oostendorp, H., Prakken, H., and Vreeswijk, G.A.W. (2006). A critical review of argument visualization tools: do users become better reasoners?. In *Working Notes of the 6th Workshop on Computational Models of Natural Argument* (CMNA2006), 67-75.

van Gelder, T. (2007). Rationale: Making people smarter through argument mapping. *Law, Probability and Risk*.

van Gelder, T. (2007). The rationale for Rationale™. *Law, Probability and Risk*, 6(1-4), 2342.

Vandergriff, I. (2006). Negotiating common ground in computer-mediated versus face-to-face discussions. *Language Learning & Technology*, 110-138.

Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research*, 342-365.

Venkatesh, V., & Davis, F.D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, 186-204.

Vassileva, J., & Sun, L. (2007). An improved design and a case study of a social visualization encouraging. *Groupware: Design, Implementation, and Use*, 72-86.

Viegas, F. B., Wattenberg, M., Van Ham, F., Kriss, J., & McKeon, M. (2007). Manyeyes: a site for visualization at internet scale. *Visualization and Computer Graphics, IEEE Transactions on*, *13*(6), 1121-1128.

Walton, D. (1996). Argumentation Schemes for Presumptive Reasoning, Lawrence Erlbaum Associates: NJ

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press.

Whittaker, S., Terveen, L., Hill, W. & Cherny, L. (1998). The dynamics of mass interaction. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*, 257-264.

Wikes-Gibbs, D. & Clark, H.H. (1992). Coordinating beliefs in conversation. Journal of Memory and Language. 31 (2), 183–194.

Wingmore, J. (1913). T*he principles of Judicial Proof*. Little Brown & Company, Boston, USA.

Yearwood, J. L., & Stranieri, A. (2006). The generic/actual argument model of practical reasoning. *Decision Support Systems*, *41*(2), 358-379.

279x164mm (96 x 96 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

279x142mm (96 x 96 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

163x130mm (96 x 96 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
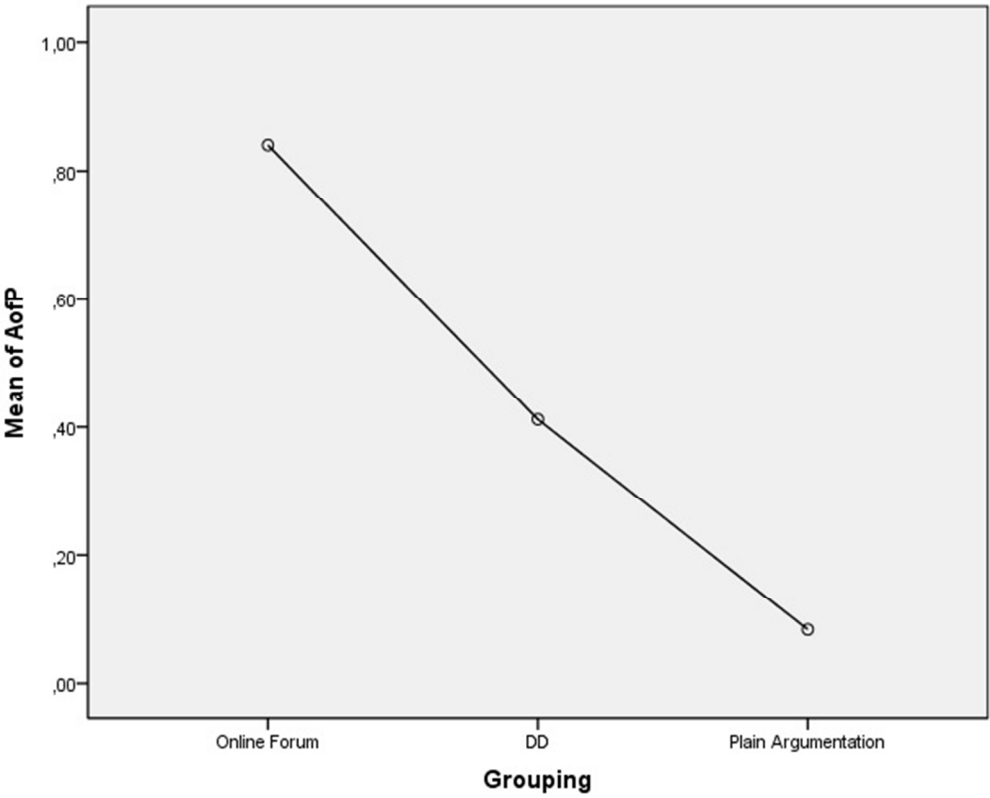58
59
60

164x129mm (96 x 96 DPI)
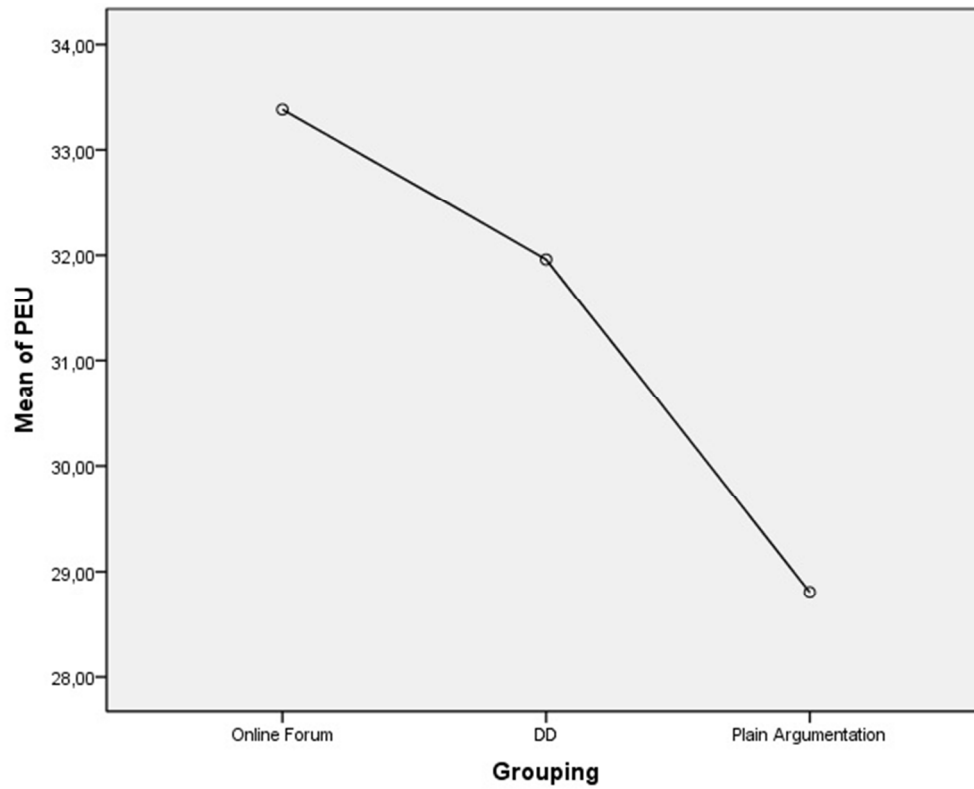
164x129mm (96 x 96 DPI)

162x129mm (96 x 96 DPI)

Table 1. Conversational affordances in mediated communication

Affordances	Clark *et al.'s* definition	Adapted to web mediated conversation
Audibility	Participants hear other people and sound in the physical environment	Participants hear other people and sound in the virtual environment
Copresence	Participants share the same physical environment	Participants are mutually aware when they are synchronously sharing the virtual environment
Cotemporality	Participant B receives at roughly the same time the message as user A produces it	Participant B receives at roughly the same time the online message as user A produces it in the virtual environment (in real time)
Mobility	Participants can move around a physical space	People can communicate their position in the shared virtual environment
Reviewability	Participant B can review A's message	Messages do not fade over time but can be reviewed
Revisability	Each participant can revise his messages	Message can be revised before being sent or edited afterward
Simultaneity	Participant can send and receive at once and simultaneously.	Participants can send and receive messages at once and simultaneously
Sequentiality	Participant A's and B's turns cannot get out of sequence.	Participants can use and see the reply structure
Tangibility	Participants can touch other people and objects in the physical environment	NA
Visibility	Participants A and B are visible to each other	Participants can see each other or at least some of the actions performed by other users in the shared virtual environment

Table 2: Cronbach's alpha

Constructs	Cronbach's α
Perceived Effectiveness of Collaboration	0.72
Mutual Understanding	0.75
Perceived ease of use	0.85

Table 3: Discriminant validity

CONSTRUCT	Sqrt AVE	SIC
Perceived Effectiveness of Collaboration	0.81	0.53; 0.22; 0.58
Mutual Understanding	0.91	0.53; 0.27; 0.79
Perceived ease of use	0.88	0.79; 0.58; 0.21

Table 4: The measurement of Nonindependence

DEPENDENT VARIABLE	INTERCLASS CORRELATION	F statistic
Perceived Effectiveness of	0.3627[*]	18.641

Collaboration		
Mutual Understanding	0.3699*	19.199
Perceived ease of use	0.1396*	6.029

*$F_{tabulated}$=3.15 and *p-value*=0.05

Table 5: Adjusted p-values for the ANOVA tests

Dependent Variables	*adjusted* F	*adjusted* Degree of Freedom *(df')*	*adjusted* p-values
MU	19.368	29.567	0.0000
PEC	18.8054	30.6602	0.0000
PEU	6.0923	91.6801	0.0032

Table 6: Significant differences among groups as whole

DEPENDENT VARIABLES	F	P-VALUES
Mutual Understanding (MU)	19.199	0.0000*
Perceived Quality of Collaboration (PEC)	18.641	0.0000*
Perceived ease of use	6.029	0.003*

*statistically significant with adjusted *p-values* (Table 5)

Table 7: Anova test results

Multiple Comparisons					
Dependent Variable	Hypotheses	(I) Grouping	(J) Grouping	Mean Difference (I-J)	P-values
MU	H1	Online Forum	DD	3,0271	0,0205
			Plain AMT	6,2303*	0,0000*
	H1a	Online Forum	DD	3,0271	0,0205
	H1b	DD	Plain AMT	3,2033	0,0120
PEC	H2	Online Forum	DD	3,0082	0,0261
			Plain AMT	6,3104*	0,0000*
	H2a	Online Forum	DD	3,0082	0,0261
	H2b	DD	Plain AMT	3,3022	0,0117
PEU	H3	Online Forum	DD	1,4224	0,6046
			Plain AMT	4,5768*	0,0029*
	H3a	Online Forum	DD	1,4224	0,6046
	H3b	DD	Plain AMT	3,1544	0,0852

*statistically significant with *adjusted p-values* (Table 5)