The Dangers of Assuming Before Analysis: Three Case Studies of Argumentation and Cognition

Kristine Lund and Matthieu Quignard
kristine.lund@ens-lyon.fr, matthieu.quignard@ens-lyon.fr
ICAR, CNRS, Ecole Normale Supérieure Lyon, University of Lyon

Abstract: In this article, we argue that researchers in the hypothetico-deductive tradition expose themselves to dangers when they first make assumptions about theoretical constructs and second, when they gather data specifically in order to test the predictions arising from hypotheses about these constructs. We review our own research in computer supported collaborative learning and in computer supported collaborative work on argumentation and cognition to show ways to partially surmount these dangers while raising new methodological concerns. We conclude by underlining the importance of the role of theory and the importance of reflecting on these issues, especially in a multidisciplinary community such as the learning sciences.

Our vision of theory may expose us to dangers

The ways in which a researcher leverages theory in the human sciences directs the types of questions asked, the type of participants who will be studied, the way data will be collected, how the data will be analyzed and how results will be presented (Kawulich, 2009). In the hypothetico-deductive framework, researchers make hypotheses about theoretical constructs and data is gathered specifically in order to test the predictions arising from hypotheses. In what follows, we review three of our own research projects, mainly situated within the hypothetico-deductive approach. Our goal is to illustrate a set of dangers that a researcher could encounter when making assumptions about theoretical constructs before analysis and to show how each danger can be partially surmounted. We hope to encourage reflection on the possible consequences of the epistemological foundations of chosen research methods.

Making theoretical assumptions about conceptual constructs prior to collecting data

The hypothetico-deductive approach in Computer Supported Collaborative Learning is illustrated by what can be called the Munich group (e.g. Stegmann & Fischer, 2011; Weinberger, Ertl, Fischer & Mandl, 2005). Theirs is a brand of research where theoretical assumptions are made before gathering the data that will be used to test them. This does not mean that theory is not also elaborated from previous empirical studies; it just means that a question is formulated before collecting the data for a particular study aimed at answering that question. In what follows, we note examples of these researchers' questions, the type of participants they study, their data collection, analysis and presentation of results and show how they are all influenced by their adopted theoretical position. The Munich group builds studies in order to test theoretical assumptions of relations between specific features of collaborative processes and successful knowledge construction. A theory used in this sense of the term is a general proposition, or a logically connected system of general propositions, which establishes a relationship between two or more variables, in general – independently of time and place (Abend, 2008).

Research questions and the notion of learning

The hypothetico-deductive method is used where hypotheses are proposed, predictions are generated and then tested in an experimental setting in order to find out to what extent the hypothesis is founded. For example, a typical question would be: "To what extent are specific epistemic activities (e.g. relating conceptual and problem space) associated with improved domain knowledge among participating individuals?" (Stegmann & Fischer, 2011). The prediction corresponding to this question is that completeness of single arguments during discussions is positively related to depth of cognitive processing of the individual and that longer argumentation sequences (those that include for example counter-arguments) are positively related to improved domain knowledge among participating individuals. Another example research question is: "What are the effects of an epistemic script and a social script and their combination on the individual acquisition of knowledge as the outcome of collaborative learning in a text-based computer-supported peer discussion environment?" Weinberger, et al. 2005). Predictions here involve both types of scripts (alone and in combination) enhancing individual knowledge acquisition in comparison with a non-scripted environment. Although these research questions are "confirmative", research questions can also be more "exploratory" in order to generate hypotheses (Stegmann & Fischer, 2011). Regarding both of the research questions mentioned above, learning is defined in terms of individual acquisition of knowledge. Generally, such acquisition is measured by differences between a pre-test and a post-test, in between which an experiment with different conditions takes place (see more below).

Participants, data, analysis and presentation of results

When researchers studying learning use an experimental method and have predictions to test or questions to explore, they often choose their participants from students already taking a class and insert their pedagogical experiment into the lesson plan. The studies referred to in Stegmann & Fischer (op. cit.) were carried out at the University level and the example data shown were collected from an online discussion (threaded chat) that used a collaboration script to prompt students to elaborate grounds and qualifications for claims they make about a particular situation and so was specifically set up to test the aforementioned predictions arising from theory and previous empirical studies.

Data were analyzed by applying measures that evaluated the quality of single arguments, the quality of argumentation sequences, and the quality of individual domain knowledge. The categories used in analyzing the quality of the collaborative process can be developed in a top down or bottom up manner. In the former, the effects and relations between the variables being examined have been well conceptualized in theory whereas in the latter, these effects and relations are to be explored during analysis. Often, there is a mix of both (Stegmann & Fischer, op. cit.). In such studies, results are presented in terms of the extent to which a particular experimental condition (analyzed in terms of the quality of their process) gave rise to a statistically significant result in terms of individual domain knowledge gains. In the case where no pre-test is possible (e.g. due to a lack of initial knowledge on the part of the participants or because the goal is to trace knowledge expressed individually during a post-text to knowledge expressed during collaboration), the post-test is evaluated alone for its expression of knowledge, using scales adapted to content.

Results are often expressed in terms of ANOVA calculations showing significant effects of experimental conditions on outcome measures. Treatment checks are sometimes also performed in order to ensure that participants reacted to the intervention as was intended by the experiment designer and these measures are also presented as results.

Summary of the hypothetico-deductive approach

This discussion of the typical experimental psychology approach to studying learning (some would say quasi-experimental as these studies are often not carried out in strict laboratory conditions, but rather in a classroom) shows that the consequences of a researcher's chosen theoretical framework are felt throughout the whole research process. The definition of learning is specific to the approach and the measurement of learning is highly dependent upon assumptions about it. In addition, this experimental psychology approach determines the way participants are chosen, data collected, analyzed and presented. The hypothetico-deductive method theorizes about the relations between conceptual constructs and their effect on outcome measures. Experimental research is accused of being less ecologically valid than, say, ethnomethodological research in addition to being "a product of the researcher's or informant's manipulation, selection, or reconstruction of preconceived notions of what is probable or important" (ten Have, p. 2), but ethnomethodological approaches are accused of not obtaining results that are as generalizable. That said, there are numerous studies that give reasons for experimental research failing to generalize, so this method is not immune to this criticism either. For example different results may be obtained when a variable that was not taken into consideration originally is paid attention to, thus illustrating the importance of context. As a case in point, McGarrigle & Donaldson (1974) showed how the experimenter crucially influenced children for Piaget's traditional conservation task. These authors described how children were taking into account the actions of the experimenter and not just the experiment, per se, when formulating their answers to the questions the experimenter asked them. In brief, if the experimenter was performing an action and then asking the children if something changed, they responded that something did change because why would the experimenter ask them if something had not changed? This taking into account theory of mind (Antaki, 2004) as an important element in interpreting experimental results was illustrated when the children responded correctly when a "naughty teddy bear" intervened, manipulated by the experimenter, and who, according to the experimenter, had the habit of "messing up toys" or "spoiling the game" (McGarrigle & Donaldson (1974). Bringing this alternative explanation of results to light gave a potential explanation for why it was difficult to replicate the age at which children were supposed to understand conservation of matter.

Clearly, experimental methods are designed with a mind to controlling conditions so that relations between constructs and outcomes can be clearly established, but it's also clear that researchers focus on particular aspects of the situation that they deem important, in light of their world view (e.g. what they consider is important to pay attention to in regards to learning), thereby allowing them to miss other influential factors. In a more basic sense, there can be problems with experimental protocols (e.g. small effect sizes, etc.), but that is not our argument here. Instead, we argue that viewing methods and results that are derived from them with different perspectives is an effective way to eliminate alternative explanations for results and to thus have more confidence in the causal link between condition/treatment and outcome. Schooler (2014) argues for Metascience — the science of science — where the goal of researchers is to examine how scientific practices may influence the validity of scientific conclusions. It is an argument both for careful methodologies and recognition of underlying assumptions and biases.

In the remainder of this paper, we take a closer look at one particular type of danger in the hypothethico-deductive approach that researchers expose themselves to: formulating theoretical assumptions about conceptual constructs prior to collecting data. In what follows, we first recount an anecdote to illustrate what pre-theorizing before analysis can lead to. We then review our own work on cognition and argumentation and show how we surmounted a set of dangers due to pre-theorization. In doing so, we are giving the backstory to the studies we refer to, something that is rarely done. Granted, analytical methods courses address some of the issues we raise, but nevertheless, some researchers do not question how their underlying epistemologies influence their work (Lund, Rosé, Suthers, & Baker, 2013). Although we do not have space to fully compare the hypothetico-deductive method with ethnomethodological approaches — two important approaches in the CSCL community — it is our hope that building a narrative regarding some of our experiences with the first approach will enable young researchers to anchor such reflections in their own practice. We conclude by underlining the importance of the role of theory and the importance of reflecting on these issues, especially in a multidisciplinary community such as the learning sciences.

Changing your perception of data to fit your theory

The following anecdote illustrates quite clearly one of the dangers of having assumptions about theoretical concepts before analysis. A researcher in particle physics was invited to the first author's home; let's call him Pierre. Upon seeing the pet cat, Pierre said: "That cat must be a female, it's a calico three-color cat". Indeed as genetic theory specifies (Kaelin & Barsh, 2013), the gene that determines how the orange color is displayed is on the X chromosome. Female cats have two X chromosomes whereas male cats have an X and a Y chromosome. A cat can only be calico if it has two X chromosomes and so in the majority of instances, the cat will be a female. However, a calico cat can be male if the cat has three sex chromosomes — two X and one Y, although this is extremely rare. When the first author told Pierre that the cat was male, he did a surprising thing. He altered his perception of the data. Instead of recognizing the three colors he originally saw, he said he instead now saw only two colors. It seems that Pierre was capable of changing what he saw in the data so that the data fit his theory; it was apparently too risky for him from a statistical point of view to imagine that he was looking at the rare case of a calico male.

This anecdote is similar to confirmation bias (Nickerson, 1998). This bias occurs when researchers actively seek evidence that confirms their hypothesis while ignoring evidence that could disconfirm it and is not limited to researchers trained in the exact sciences (e.g. particle physics). In the case of Pierre, even though he saw one thing initially, he changed his perception to be more compatible with what he should have been seeing, according to the theory. In fact, he distorted his view of the data (albeit presumably not intentionally) so that it would be in alignment with the expectations of the pre-chosen theory. Although such behavior can be considered to be a danger to good scientific inquiry, many researchers risk such behavior as choosing a theory that makes predictions about data before gathering data is a popular way of doing research — in the exact sciences, but also in the human and social sciences. In the next sections, we describe three of our own studies where we were exposed to the dangers of making assumptions about theoretical constructs before data analysis. Each study illustrates a particular danger and describes how we surmounted it.

The study of cognition and argumentation

As with any domain of study, researchers interested in cognition and argumentation take different theoretical approaches. In this section, we concentrate on research falling mainly into the hypothetico-deductive approach, where conceptual constructs of argumentation are defined according to theory and then studied in computer-supported interaction. These conceptual constructs were assumed to predict learning gains in the first example, built into patterns that illustrated procedures of decision-making in the second example and used to predict the amount of conflict in Wikipedia in the third example.

Widen the context of the phenomenon of interest

The danger of making assumptions about theoretical constructs illustrated by our first example is potentially not as serious as confirmation bias, but it is one where the researcher's focus may be considered to be too narrow and therefore has the consequence of excluding other phenomena that may also have repercussions for the outcome measures. As Greeno (1998) points out, if we investigate cognitive subsystems, some general activity structures have to be arranged in which these subsystems function. It is currently the case that we don't understand the relations between the different subsystems, yet when we run an experiment, we are still assuming that the particular subsystem we are investigating does not significantly depend on how the other subsystems behave. This is most likely not the case and so we risk generalizing in an unjustified manner.

This research originated within the European project SCALE where we elaborated a coding scheme called Rainbow, for analyzing on-line pedagogical debates during which students could both chat and draw argument diagrams (Baker, Andriessen, Lund, Amesvoort & Quignard, 2007). The scheme was inspired both by theory (e.g. Toulmin, 1958; Barth & Krabbe, 1982; van Eemeren & Grootendorst, 1984; Plantin, 1990) and other coding schemes (e.g. Meier et al. 2007; Suthers, 2006; Andriessen, Erkens, van de Laak, Peters, & Coirier, 2003; van Bruggen and Kirschner, 2003 and Veerman, 2003; de Vries, Lund, & Baker, 2003) and further elaborated in confrontation with part of our own corpus. It was then applied to the rest of our corpus and validated with inter-coder reliability. We obtained results relating specific on-line distance pedagogical situations (e.g. using an argumentation diagram as a means of debate or using it as a means for representing a debate) and quality of an outcome measure of argumentation (e.g. Lund, Molinari, Séjourné & Baker, 2007), but we were also victims of our own pre-conceived view of what was important, according to our dominant theoretical framework. For example, we attributed the most importance to categories 5-7 (cf. Table 1) because that was where the conceptual notions under debate were being delineated, disentangled and deepened and we were interested in how such processes related to learning.

Table 1: The Rainbow coding scheme initially elaborated for analyzing pedagogical debates (cf. Baker, Andriessen, Lund, Amesvoort & Quignard, 2007)

1) Outside Activity	Any interaction that is not concerned with interacting in order to carry out the teacher/ researcher-defined task, including socio-relational interaction that does not relate to interacting in order to achieve the task, e.g., talk about last night's party.
2) Social Relation	Interaction that is concerned with managing the students' social relations with respect to the task (debating about X), e.g. greeting, leave-taking, politeness, expressions of frustration with the way the partner is interacting, etc.
3) Interaction management	Interaction concerned with managing the interaction itself: who will speak or not and when (coordination), establishing contact, perception, understanding, attitudes (communication management), topic shifting, time management, ...
4) Task management	Management of the progression of the task itself: planning what is to be discussed, establishing whether problem solved or not, ...
5) Opinions	Interaction concerned with expressing (stating, requesting) opinions (beliefs, acceptances, ...) with respect to the topic debated, especially (but not only) at opening and closing of sequences of argumentative discussion (dialectical outcomes).
6) Argumentation	Expression of (counter-)arguments directly related to a thesis (e.g. GMOs increase famine because farmers become dependent on seed companies), theses themselves, requests for justification
7) Explore and deepen	Interaction concerned with (counter-)arguments linked to (counter-)arguments, argumentative relations, and meaning of arguments themselves (elaboration of them, definition, extension, contraction, i.e. any discursive or conceptual operation performed on content of arguments themselves).

However, other types of interaction on which we were not immediately focused turned out to also be important for the way argumentation unfolded. For example, the dynamics of power and influence between group members can explain to what extent some members admit publicly to being influenced by other members in how they change their opinions (Molinari & Lund, 2012). In Baker, Andriessen & Lund, (2009), we explore extending the socio-cognitive paradigm to include how the dynamic interplay of emotions relate to processes of knowledge co-elaboration (see also Baker, Andriessen & Järvelä, 2013 and Polo, Lund, Plantin, & Niccolai, 2016). Finally, even a phenomenon specifically judged as initially uninteresting by the theoretical framework can be revealed to be interesting later on. For example, although spatially arranging elements of an argumentation graph might have been thought to be a waste of time, as opposed to arranging them logically, or thematically, it can also allow students to review the arguments discussed during the interaction and therefore increase the quality of their argumentative texts written afterwards (Baker, Quignard, Lund & Séjourné, 2003).

These examples illustrate that although we initially focused on aspects that restricted our understanding of the phenomena that interested us, we were subsequently able to explore other aspects that allowed us to more fully explain the phenomena; the question now focuses on combining these insights. This is not an illustration of confirmation bias, per se, where researchers only see evidence that support the conclusions they want to make.

But it still shows how suppositions about theoretical constructs orient the gaze of researchers. The first studies may be represented by the old adage of looking for your keys where the light is good (e.g. first we looked at conceptual aspects of argumentation) whereas during the further analyses, we turned on other lights (e.g. highlighting social relations and emotion during argumentation and spatial organization of arguments) and looked under *those*. The idea is to link up all the different light sources and build a coherent narrative.

Code not to count, but to render phenomena observable

Our second example comes from a long term project focused on studying decision making during collaborative design in industrial engineering contexts (e.g. Lund, Prudhomme & Cassier, 2013; Prudhomme, Pourroy, Lund, 2007). We began this study in a similar way to the initial studies carried out within the SCALE project. Inspired by theory (Plantin, 2016; Simon, 1969; Gero, 2002; Vera, 2003; Hutchins, 2000) and previous empirical results (e.g. Baker et al. 2007, Détienne, Boujut & Hohmann, 2004), we developed an initial coding scheme (cf. Figure 1) and further elaborated it in confrontation with part of our corpus, using the rest of the corpus to perform inter-coder reliability. We did yet not have specific predictions to test by a coding and counting procedure. We were using the hypothetico-deductive approach but in an initial more exploratory and hypothesis generating stage. That said, our theoretical bases and empirical experience directly underpinned the nature of our coding scheme but the scheme's initial set-up was unable to show us the *criteria* that designers use during decision-making, when they argue in favor or against solutions.

Pragmatic functions of Interaction	**Subjects of Interaction**
Management	Project
Proposition	Task
Explanation	Solution
Argumentation	Criteria
Opinion	Tool
	Social Relation
	Communication

Figure 1. The Design Interaction Framework (DIF).

Our coding system enabled us to see how designers employed the pragmatic functions of their utterances (e.g. for managing, or for arguing) on particular subjects of interaction (the project, the solution). The arrows show two possible ways – a designer can manage a project and argue about a solution and argue about a solution. But he or she can also manage a task, a tool, a social relation, or communication. And he or she can argue about the project, per se, about a task, solution, criteria, tool, social relation or communication. We did not find all possible combinations of pragmatic functions of the interaction applied to subjects of the interaction in our data — for example, we did not find designers who managed solutions or criteria.

Although our coding system enabled us to make sense of how the design process evolved, it did not allow us to pinpoint exactly how designers used criteria when they argued. In other words, when argumentation was carried out on proposed solutions, the criteria each designer used to evaluate the solution were mobilized within the arguments and were thus not displayed with the way the coding scheme was set up. We could see arguments *about* criteria, but not arguments about solutions that *mobilized* criteria. So we went back to the corpus and specifically looked at how criteria were mobilized during argumentation. This work enabled us to build static and dynamic visualizations of how criteria are mobilized in collaborative design and illustrate two patterns of decision-making we found multiple times, in two different design contexts. The first pattern is when the criteria mobilized allow designers to choose between solutions already on the table and the second is when the criteria mobilized force the designers to propose and choose a new solution.

We argue that this example illustrates a way to potentially escape some of the limits of coding, often criticized in the literature: coding and counting cannot adequately capture important aspects of human interaction such as emotion (Peräkylä, 2004) and counting removes utterances from their meaning-making context (Schegloff, 1993). Here, we used coding as an exploratory aid in understanding the detailed articulation between argumentation and decision-making. In particular, systematically studying how criteria specific to a particular profession (e.g. mechanical engineering, assembly line work, project managing) were mobilized during arguments for and against solutions that were proposed during collaborative design allowed us to bring to light two patterns of decision procedures, without any counting of categories. In other words, coding can be seen as a filter that brings to light a phenomenon that could not have been noticed had the coding not been done, thus enabling a detailed understanding of the interactive mechanisms.

Admit that even simple indicators can have explanatory power

In our last example, we show how comparing methods for predicting the existence of conflict in Wikipedia forums enabled us to 1) become aware of the validity of a model based on theory we would not have initially considered and 2) to evaluate the relative contributions of a model built from data versus one built on theory (Denis, Quignard, Fréard, Détienne, Baker, & Barcellini, 2012).

We began by elaborating an automated tool, based on natural language processing (NLP) techniques that could categorize Wikipedia threads as containing (or not) conflicts. Our tool was elaborated on theory stemming from Barth & Krabbe (1982) and Mackenzie, (1985) where a dispute can be modeled as a dialogue game whose goal is to attack or defend a controversial statement and eventually solve the conflict. According to Barth & Krabbe (op. cit.) a conflict is formally defined when participants have committed to at least one attack and one defense move with respect to a statement. By those moves, participants overtly take position for and against a point of view and commit to their respective positions as long as they can. Such a dialectical model gives us a practical method for identifying conflicts: find two utterances with opposite argumentative orientations (for or against) in relation to a preceding third utterance. However, implementing this method with NLP techniques is not trivial and involves both checking for markers of first and second person (e.g. me, you, your, etc. to distinguish viewpoints) and for the global connotation of the utterance (to obtain positive or negative nature). According to our method, there was a conflict if there were at least two negatively connoted contributions, one with a first person marker and the other with a second person marker. For a corpus of 320 discussions (122 were in conflict, or 38%), the method succeeds 77,8% of the time (0.86 of f-measure, kappa of 0.7), as compared to an expert analysis.

We compared this result to another method applied to the same corpus in order to appreciate its efficiency and to test whether or not the same markers used with simpler decision rules gave better results. The statistical induction method (Quinlan, 1993) calculates the global rates for each marker across the corpus and finds the most pertinent thresholds. Contrary to our own decision rule, this one is very simple to implement: if the discussion is at least 5 messages and if there are more than 8 2nd person markers per 1000 words, then there is a good chance that the text contains conflicts. This simpler rule has a success rate of 50% (0.64 of f-mesure, kappa of 0.5), compared to an analysis done by an expert. Although this is a better result than by chance (which would be 38%) and significantly worse than our own, it still is surprisingly accurate, for its simplicity.

It is impossible to predict beforehand what type of rule will result from this type of induction method. What is surprising is that this rule does not predict conflicts as much as it predicts the co-existence of multiple voices that are interacting with each other. And since there is an overlap between the co-existence of two voices and expressed conflicts, this rule also predicts conflicts, but less efficiently. We can therefore draw two main lessons from our experience. First, the second rule made us aware of a global characteristic of the forums (i.e. presence of voices that interact); this could partially explain a local characteristic (i.e. presence of conflict). Remember that forums do not necessarily have multiple voices. There can be only one voice in the "discussion" area that never interacts with anyone else. Focusing on the interacting multiple voices forced us to consider the importance of this second rule, something we would never have done as the assumptions we made about the theoretical constructs involved oriented us to thinking that this rule would be unimportant. Second, this result also forced us reconsider the contribution of statistical approaches (driven by data) in relation to symbolic approaches (driven by theoretically based models). The former allows us to test the explanatory power of the markers without presuming how they should be interpreted whereas the latter furnishes the way in which the markers should be interpreted. This result argues for alternating data driven and theoretically motivated approaches.

Conclusions

In this paper, we have described the hypothetico-deductive approach for carrying out research oriented to cognition and argumentation in computer supported collaborative learning and in computer supported work and problem solving situations. We have discussed some of the theoretical assumptions mobilized in this approach, how theory is used and its influence when it comes to defining research questions, conceptualizing learning, gathering and analyzing data, and representing results in relation to cognition and argumentation. We chose to examine one of the weaknesses of the hypothetico-deductive method — that of the different dangers researchers expose themselves to, due to formulating theoretical assumptions about data constructs prior to collecting data. In examining three of our own research projects, we illustrated three specific dangers due to making such assumptions and how we were able to partially surmount each of them, while also posing new methodological questions. First, the design of coding and counting schemes elaborated to test predictions of relations between conceptual constructs concerning collaborative processes and individual learning gains necessarily orient the gaze of the researcher toward particular phenomena. The trick is to be able to switch foci and break free of one's first assumptions. If a researcher can focus her gaze on other phenomena, this can lead to the discovery of other significant relations. But in this case, one must also build a narrative that coherently combines all of these results. The question then arises as to how inclusive a narrative should be before it is deemed sufficient. Second, elaborating a coding and counting scheme for fine grained exploratory analysis of collaborative processes also orients a researcher's gaze, but avoiding specific initial hypothesis testing and avoiding counting codes opens up the analyst to attuning to what can emerge from the corpus, while keeping the interactional context intact. But in this case, how should generalization be addressed? We answered this question by seeking and finding the same patterns of decision-making in two very different design contexts (Prudhomme, Pourroy, & Lund, 2007); Lund, Prudhomme, & Cassier, 2013), thus adding strength to an argument for generalization. Finally, comparing results in a benchmarking context where each result stemmed from a different method forces the researcher to question the assumptions underlying her method and to recognize the intangibility of her theoretical framework. Here, we wonder how we may place theoretically motivated and data driven approaches in dialogue to one another.

Our three examples are particularly targeted toward young researchers who may not have begun to reflect on the epistemological foundations of their research methods, or on how their practices may influence both the focus of their work and the validity of their results. This is more common than may be supposed. Finally, in a community such as the learning sciences where researchers collaborate in multi-disciplinary contexts, it's important to consider the theoretical assumptions that underlie our research. The epistemological encounters that researchers from different traditions may experience may be leveraged in order to explore the extent to which approaches can be integrated (Lund, Rosé, Suthers, & Baker, 2013).

References

Abend, G. (2008). The Meaning of 'Theory'. *Sociological Theory*. 26(2), 173-199

Antaki, C. (2004). Reading Minds or Dealing with Interactional Implications? *Theory & Psychology*, Sage Publications, 14(5): 667–683.

Baker, M.J., Andriessen, J., Lund, K., van Amelsvoort, M., & Quignard, M. (2007). Rainbow: a framework for analysing computer-mediated pedagogical debates. *International Journal of Computer-Supported Collaborative Learning*. 2:315–357.

Baker, M.J., Quignard, M., Lund, K. & Séjourné, A. (2003). Computer-supported collaborative learning in the space of debate. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds.) Designing for Change in Networked Learning Environments : *Proceedings of the International Conference on Computer Support for Collaborative Learning 2003*, pp. 11-20. Dordrecht : Kluwer Academic Publishers.

Baker, M., Andriessen, J., Lund, K. (2009). Socio-relational, affective and cognitive dimensions of CSCL interactions: integrating theoretical-methodological perspectives. Symposium in the proceedings of the *International Conference of Computer Supported Collaborative Learning (CSCL09) on CD-ROM*. June 13-18, University of the Aegan: Rhodes, Greece.

Baker, M.J., Andriessen, J. & Järvelä, S. (2013). *Affective Learning Together, Social and emotional dimensions of collaborative learning*. NewYork: Routledge.

Barth, E. M., & Krabbe, E. C. W. (1982). *From axiom to dialogue: A philosophical study of logics and argumentation.* Berlin: Walter de Gruyter.

Denis, A., Quignard, M. Fréard, D., Détienne, F., Baker, M & Barcellini, F. (2012). Détection de conflits dans les communautés épistémiques en ligne. Dans Conférence Traitement Automatique des Langues Naturelles, Grenoble, 2012.

Détienne, F., Boujut, J.-F., & Hohmann, B., (2004). Characterization of collaborative design and interaction management activities in a distant engineering design situation. In (Eds.) F. Darses, R. Dieng, C. Simone & M. Zacklad, *Cooperative systems design,* (pp. 83- 98). Amsterdam, The Netherlands: IOS Press.

de Vries, E., Lund, K. & Baker, M.J. (2002). Computer-mediated epistemic dialogue: Explanation and argumentation as vehicles for understanding scientific notions. *The Journal of the Learning Sciences*, 11(1), 63-103

van Eemeren, F. H., & Grootendorst, R. (1984). *Speech acts in argumentative discussions*. Dordrecht: Foris.

Gero, JS (2002) Computational models of creative designing based on situated cognition, in T Hewett and T Kavanagh (eds), *Creativity and Cognition* 2002, ACM Press, New York, NY, pp. 3-10.

Greeno, J. G. (1998). "The situativity of knowing, learning, and research". *American Psychologist* 53 (1): 5–26.

Hutchins, E. (2000). Cognition in the Wild. MIT Press, London.

Kaelin, C.B. & Barsh, G. S. (2013) Genetics of Pigmentation in Dogs and Cats. Annual Review of Animal Biosciences. Vol. 1, 125-156

Kawulich, B. (2009). The Role of Theory in Research. In Garner, M., Wagner, C., & Kawulich, B. (Eds.). *Teaching research methods in the social sciences* (pp. 37-82). Burlington, VT: Ashgate.
Lund, K., Molinari, G., Séjourné, A. & Baker, M.J. (2007). How do argumentation diagrams compare when student pairs use them as a means for debate or as a tool for representing debate?". *International Journal of Computer-Supported Collaborative Learning*. 2:273–295.
Lund, K., Rosé, C. P., Suthers, D. D., & Baker, M. (2013). Epistemological encounters in multivocal settings. In D. D. Suthers, K. Lund, C. P. Rosé, C. Teplovs & N. Law (Eds.), *Productive Multivocality in the Analysis of Group Interactions.* In C. Hoadley & N. Miyake (Series Eds.), *Computer Supported Collaborative Learning Series*: Vol. 15 (pp. 659-682). New York: Springer.
Lund, K., Prudhomme, G., & Cassier, J.L. (2013). Pivotal moments for decision making in collaborative design: are they teachable? In (Eds.) S. Goggins, I. Jahnke & V. Wulf. CSCL@Work: Case Studies of Collaborative Learning at Work (Computer-Supported Collaborative Learning Series) (pp 243-268). New York: Springer.
Lund, K., Rosé, C. P., Suthers, D. D., & Baker, M. (2013). Epistemological encounters in multivocal settings. In D. D. Suthers, K. Lund, C. P. Rosé, C. Teplovs & N. Law (Eds.), *Productive Multivocality in the Analysis of Group Interactions.* In C. Hoadley & N. Miyake (Series Eds.), *Computer Supported Collaborative Learning Series*: Vol. 15 (pp. 659-682). New York: Springer.
Mackenzie, J. D. (1985). No Logic before Friday. *Synthese*, 63:329–341.
McGarrigle & Donaldson (1974). Conservation Accidents. *Cognition*. 3(4), 341-350.
Molinari, G. & Lund, K. (2012). How a power game shapes expressing opinions in a chat and in an argument graph during a debate: A case study. In J. van Aalst, B.J. Reiser, C. Hmelo-Silver, K. Thompson (Eds.), The Future of Learning*: Proceedings of the 10th International Conference of the* Learning Sciences (ICLS 2012), *(Vol. II, pp. 232-236).* July 2-6, Sydney: International Society of the Learning Sciences.
Nickerson, R. S. (1998). Confirmation Bias; A Ubiquitous Phenomenon in Many Guises, *Review of General Psychology* (Educational Publishing Foundation) 2(2), 175–220.
Peräkylä, A. (2004). Two traditions of interaction research. *British Journal of Social Psychology*, 43, 1-20.
Plantin, C. (1990) *Essais sur l'argumentation.* Paris : Kimé.
Plantin, C. (2016). *Dictionnaire de l'argumentation – une introduction notionnelle aux etudes d'argumentation*. Lyon: ENS Editions.
Polo, C., Lund, K., Plantin, C. & Niccolai, G. (2016). Group Emotions: The Social and Cognitive Functions of Emotions in Argumentation. *International Journal of Computer Supported Collaborative Learning.* 11, 123–156.
Prudhomme, G., Pourroy, F., & Lund, K. (2007). An empirical study of engineering knowledge dynamics in a design situation. *Journal of Design Research*. 6(3), 333-358.
Quinlan, J. R. (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers.
Schegloff, E.A. (1993). Reflections on Quantification in the Study of Conversation. *Research on Language and Social Interaction*, 26(1), 99-128.
Schooler, J. W. (2014). Metascience could rescue the 'replication crisis'. *Nature*. 515 (7525): 9
Simon, H. (1069). *The Sciences of the Artificial.* MIT Press, Cambridge, Mass, 1st edition
Stegmann, K., Fischer, F. (2011). Quantifying Qualities in Collaborative Knowledge Construction: The Analysis of Online Discussions. In Puntambekar, Erkens and Hmelo-Silver (Eds.) *Analyzing Collaborative Interactions in CSCL: Methods, Approaches and Issues* (pp. 247–268). NY: Springer.
Suthers, D. (2006). A qualitative analysis of collaborative knowledge construction through shared representations. *Research and Practice in Technology Enhanced Learning*, 1(2), 1–28.
Toulmin, S. E. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
Vera, A.H. (2003). By the Seat of Our Pants: The Evolution of Research on Cognition and Action. *The Journal of the Learning Sciences* 12(2), 279–284.
Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science*, 33(1), 1–30.

Acknowledgments

Research discussed here was funded by the EU (IST-1999-10664-SCALE project), the Rhône Alpes region (PhD stipend), and the French ministry (ANR-08-CORD-004 CCCP-Prosodie project), The authors are also grateful to the ASLAN project (ANR-10-LABX-0081) of Université de Lyon, for its financial support within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) of the French government operated by the National Research Agency (ANR). We would also like to acknowledge the colleagues who worked with us on the projects we review here. In particular we owe a great debt to Michael Baker.