From Mutual Diagnosis to Collaboration Engines: Some Technical Aspects of Distributed Cognition

P. Dillenbourg
TECFA (Educational Technology Unit)
School of Education and Psychology
University of Geneva
9, Route de Drize, CH1227 Carouge (Switzerland)
pdillen@divsun.unige.ch http://tecfa.unige.ch

Abstract This contribution is based on the development and the evaluation of two learning environments in which the learner had to collaborate with the machine. This experience revealed that existing knowledge-based techniques are not appropriate to support 'real' collaboration, i.e. to cover a range of flexible, opportunistic and robust interactions which enable to agents to build a shared understanding of a problem. I argue for the design of collaboration engines which integrate dialogue models at the rule instantiation level. The role of dialogue models is not simply to improve the interface. The challenge is to develop models which account for the role of dialogues in problem solving and learning. Such models would reflect current theories on 'distributed cognition', one of the approaches placed under the 'situated cognition' umbrella. Most of the implications of these theories to the design of interactive learning environments (ILEs) that have been discussed so far concern the choice of methods (e.g. apprenticeship - Newman, 1989 - or project-oriented group work - Goldman et al, 1994) or software engineering (e.g. participatory design - Clancey, 1993). I address here the implications of these theories at a more technical level. I structured the argument as a discussion with myself, which is quite natural within the distributed cognition approach where interacting agents are viewed as forming a single cognitive system.

- *Hi, nice to see you here! What did you do over these last years ?*

- Some work on collaboration... In PEOPLE POWER, the human learner collaborates with another learner played by the machine (Dillenbourg & Self, 1992). In MEMOLAB, the learner collaborates with a machine expert (Dillenbourg, Mendelsohn & Schneider, 1994).

- *Did you try them ? What did you learn?*

- My conclusion is that existing knowledge-based techniques are not appropriate to support real collaboration, that we should do basic research to develop proper 'collaboration' engines...

- *What is 'real' collaboration?*

- I mean that the machine must engage the kind of flexible, opportunistic and robust interactions which enable two agents to build a shared understanding of a problem

- *Give me an example, something which was wrong with your systems?*

- When the expert in MEMOLAB misunderstood the learner, the latter could not repair this misunderstanding. For instance, a learner said about the machine expert "He supposes that I wanted the subjects to do something during 40 seconds. I wanted the subjects to do nothing!"... In everyday conversation we permanently check that our interlocutor has understood what we meant, at least well enough to carry on the task, and we have many resources for repairing such misunderstanding (Hirst et al, 1994). This process of grounding (Clark & Marshall, 1981; Clark & Brennan, 1991), this game of mutual diagnosis and repair, is what makes collaborative learning efficient.

- *What do you mean by 'mutual' diagnosis? Diagnosis has always been implemented as a one-way process!*

- *For instance, if we talk about a weather forecast, I will progressively find out what you know about it, and reciprocally, you will catch on what I know about this topic. Moreover, I will probably find out what you think that I believe about the weather forecast and vice-versa.o, your point is that there exists a second order of diagnosis, a diagnosis of a diagnosis. But, don't you think we have enough difficulties in implementing a first level diagnosis? Considering two nested levels of diagnosis is simply unrealistic.*

- Well, one could also claim the opposite. What may be unrealistic is to expect an agent to build in one-shot a correct model of his partner. Human-human and human-machine dialogues are inherently ambiguous. The key difference is that human-human communication relies on various resources to detect and repair communication failures, which are generally lacking from human-machine interactions (Suchman, 1987). Douglas (1991) observed that students often aid the human tutor in repairing misdiagnosis. It may be more sensible to give the system the ability to incrementally and interactively build some representation of the user.

- *This is not really a new idea. Several scholars have suggested making diagnosis more interactive, since IDEBUGGY (Burton, 1982) or the idea of an inspectable learner model (Self, 1988), recently turned into collaborative modelling (Bull, Pain & Bra, 1993).*

- That's right, this idea came out of system design and experimentation, before being systematised within the situated cognition framework (Dillenbourg & Schneider, 1993). It is also based on the work of belief systems in learner modelling (Self, 1992): to model second order diagnosis you need to represent knowledge specific to an agent .

- *So, if I understand you right, what you say is that the core diagnosis process, I mean the algorithm which searches for a variation of the correct rulebase which produces the same answers as the learner, should be integrated into some kind of regulation loop.*

- That's one possibility. This regulation loop would be a dialogue structure in which the learner assesses the output of what you called 'core diagnosis' and repairs it directly (by telling) or indirectly (by initiating a new diagnosis cycle). It could use some of the computational models for dialogue repair and grounding have been recently developed (Cawsey, 1991; Traum 1994, Hirst et al, 1994). In addition, their use in student modelling implies that the system is able to reason about how it produced a misdiagnosis. Hence the usefulness of the work on formalising the learner modelling process (Cialdea, 1991; Self, 1989).

- *But, grounding does not occur in a vacuum. It uses many external referents.*

- That's right... When we used MEMOLAB with two human subjects, we had to clean the screen after each experiment because the subjects most often speak with their finger on the screen... at least the one which has not the mouse in hands! The role of screen displays in supporting conversation has been emphasised by Roschelle (1990).

- *But, in the case of human-computer interaction, the machine diagnoser does not 'see' what is on the screen.*

- We developed an inference engine which discriminates internal objects from displayed objects (Dillenbourg et al, 1994). The problem is that all relevant features of any displayed object must be explicitly defined in advance. Gestures can themselves be ambiguous. Therefore, Hirst and his colleagues (1994) encoded the object salient properties in a hierarchy in order to guess which feature an agent refers to.

- *I see another problem. This 'diagnosis regulation loop' could imply rather long diagnostic dialogues.*

- Long dialogues might be justified when the decision taken on the basis of this diagnosis is important, i.e. concerns a long sequence of interactions.

- *The trouble is that the longer the diagnosis, the higher the chances that the learner changes his beliefs over time. Then diagnosis is no longer a neutral observation process which provides the system with the information necessary to make pedagogical decisions. If diagnosis becomes more interactive, it becomes less neutral. The interactions conducted to clarify what the learner believes will tend to affect what she believes. Douglas (1991) observed that the repair of tutoring failures creates "local curriculum sequence differences".*

- I agree, but this side-effect of diagnosis regulation dialogues is not an undesirable event. It corresponds to the 'appropriation' mechanism, which is central to the socio-cultural theory: the learner learns by watching how his actions are interpreted within the expert understanding of the task (Newman, 1989). As Pea (1993a) stated: "Through interpretations by others, you may come to mean more than you thought you did" (p. 270). This appropriation mechanism plays a central role in the apprenticeship method (Newman, Griffin & Cole, 1989; Rogoff, 1991). Actually, the interpretation by another does not necessarily have to be formulated explicitly ("I think you believe this..."). It can be implicitly conveyed by the partner's next action (i.e. the system's attempt to integrate the learner's action in its strategy) and hence also applies to non-verbal interactions.

- *So finally, this idea of mutual diagnosis goes beyond the diagnosis module. It questions the distinction between diagnosis and feedback, between diagnosis and intervention. Actually, it questions the basic architecture of ILEs.*

- This is not surprising. That architecture is based on the traditional cognitive science view of cognition being bound to individuals. This is why I attempt to reconceptualize MEMOLAB from a 'distributed cognition' perspective (Pea, 1993b) which....

- *Oooo, no... 'distributed'... one of these new buzzwords... Next, you will tell me that the best method is collaborative learning.*

- Not exactly, we know collaborative learning is not always efficient. No, the 'distributed' perspective on ILEs is rather to consider that the main functions of an ILE are intrinsically collaborative. Diagnosis is collaborative...

- *You said that before...*

- ... yes, but explanation is also collaborative. An explanation is not simply built by one agent an delivered to the other, but jointly constructed (Baker, 1992). Problem solving is collaborative, provided that the expert and the learner seek to build a mutual understanding of the problem. Baker (1994) even describes teaching as collaborative...

- *...like in the study by Douglas you previously referred to, where the learner helps his tutor to repair tutoring failures?*

- Exactly. Or in the work of Fox (1991), who observes that the learner and the tutor collaborate to avoid incorrect answers from the student. In some cases, the student shows signs that she cannot answer the question and thereby invites the tutor to provide resources to answer (sometimes another question). In other cases, the tutor provides early signs of negative feed-back (e.g. "predisagreement" silence), after which the student often reformulates his answer as a question, thereby inviting the tutor to help him.

- *Ok, we agree that these main functions, diagnosis, explanation or tutoring, are accomplished jointly. Is this what you mean by 'distributed'?*

- Right. Simply stated, these functions are not viewed as the performance of an individual, but as accomplished by the group itself in order to maintain its consistency. The learner and the ILE form a single cognitive system and the system-maintenance functions are distributed over the interacting agents.

- *The same word is used in 'distributed AI' (Gasser, 1991). Basically, the task is split among subtasks which are distributed among various agents with different specific skills. I can see why such a system may work better, but I don't see how such a division of labour may lead to cognitive benefits for individuals.*

- We could argue on that point, but you are right that the term 'distributed' is a bit misleading. The cognitive effects are not due to the division of labour in itself, but to the fact that, despite the division of labour, despite the heterogeneity of skills or viewpoints, the agents succeed in establishing mutual understanding. For instance, Rogoff (1991) showed that internalization did not occur when one peer dominated decision making or did not involve the other in the decision process specific to his strategy, or in other words, that they were working as two independent cognitive systems.

- *But, look at studies on human-computer interaction... When they refer to 'joint cognitive systems' (Woods & Roth, 1988), they seek a task distribution that optimises the specific skills of humans and machines... This is really 'splitting'.*

- Right, but they also use the concept of 'cognitive coupling' (Dalal & Kasper, 1994), to emphasize the necessary congruence between the agents' features such as goals, strategies, knowledge and cognitive skills.

- *So, you mean that, in a joint system, agents must have complementary features?*

- I would rather refer to a 'joint system' when the agents build a shared understanding of the problem (Roschelle & Teasley, 1995). This shared understanding may depends less upon the agents' initial features than upon the quality of interactions. It would not be a bad summary of the studies on the efficiency of collaborative learning (Webb, 1989, 1991; Dillenbourg et al, to appear) to say that collaborative learning is efficient when the learners form a '<u>joint despite distributed</u>' cognitive system.

- *I think that 'joint despite distributed' is what Resnick (1991) means by 'shared' cognition. But, sorry to be materialist, do you mean that we need grounding mechanisms between a rule-based expert and the user?*

- It depends on what you have in mind. If you think about an expert module which produces decisions which are then delivered via a dialogue module, you precisely miss the idea of shared cognition. <u>The issue is not simply to improve the interface, but to understand how interactions affect cognition</u>. This issue will not be tackled by dissociating these two components. The basic principle is precisely that if interactions are internalised, i.e. if some traces of interactions influence problem solving, it implies that the set of problem solving operators has some isomorphism with the set of dialogue operators.

- *Sorry, I don't see exactly what you mean.*

- For instance, in PEOPLE POWER, the inference engine was implemented as a dialogue between the agent and himself: the agent attempted to agree or refute his own arguments (within some cognitive load boundaries), according to the previous arguments with its partner. In this case, the 'dialogue/problem-solving isomorphism' was obtained by reducing reasoning and dialogue operators to the most rudimentary subset, "agree" and "refute".

- *Grounding implies a much richer set of operators, such as acknowledge, repair, request for acknowledgement, ….*

- Right, <u>the challenge is to design inference engines which integrate these dialogue operators</u>. Some dialogue acts could be translated into changing the scope of a rule, or the scope of a variable. In MEMOLAB, we used an object-oriented production system (Dillenbourg et al, 1994) in which each variable was associated with a reference-class which determines the set of objects which could instantiate this variable. The scope of each variable was determined statically, when the rule was written. But, in human-human dialogue, we saw that the scope was negotiated dynamically. For instance, one learner said "It (the material) should be the same words", but the other answered "... no, the same number of syllables.", i.e. he generalised the variable 'material' (different lists of words can have the same average number of syllables). Conversely, in another dialogue, one learner said "Which material do you want?" and her partner answered "You mean the words?", i.e. she specialised the variable material (a material can be a list of various things). This negotiation is often supported by deictic gestures which help to relate general variables expressed in utterances with specific screen objects. For instance, one learner says "Next, we need to do a restitution task" and her partner clicks on 'free recall' in the list of tasks and says "Is that it?"

- *OK, let me see if I understand... In the systems that you tested, the negotiation was at the rule level: the agents used "refute" or "agree" with regard to the last fired rule. It is a 'step-by-step' interaction... And in the observed dialogues, negotiation occurs below the rule level, at the variable level, where one can use a richer set of operators such as "generalise", "specialise" or "refine".*

- Right... dialogue operators in a large sense, i.e. including gestures. The concept of 'socially distributed production' (Roschelle & Teasley, 1994) already introduces dialogue turn taking between the condition and the conclusion parts of a rule (subject 1 says "if ...", and subject-2 completes with "then ...."). My claim is that negotiation may even occur at a lower level, when the variables appearing in the conditions are matched against the problem data.

- *But, what you call the 'rule level' is arbitrary. Could you not simply do that by decomposing further the ruleset?*

- I think negotiation mechanisms should be included in the unification mechanisms. For instance, even before you speak, or at least after a few words, your listener has expectations regarding the rest of your utterance. These expectations govern his interpretations of what will follows. It even occurs sometimes that the listener finishes the sentence before the speaker. Engel and Haakma (1993) model this mechanism by the fact that the interpretation of the beginning of the message leads to select the lexicon used to parse the remaining of the message. This idea could be transposed in terms of instanciation mechanisms: the scope of a variable Z, i.e. its reference-class, could be determined dynamically on the basis of the other variables X, Y, ... instantiated in the condition clauses already matched or already discussed, in the same or in a previous rules (for instance as a link between the meta-class of the reference-classes).

- *But, this 'expectation' effect may produce misunderstandings!*

- Of course! I never said that we have to avoid misunderstanding. We simply cannot! Instead, we have to support repair mechanisms, not simply because these mechanisms play a role in understanding, but because they play a role in problem solving and learning. The 'linguistic' mechanisms by which we monitor mutual understanding cannot be different from those of involved in mutual regulation. And we know mutual regulation is central to collaborative learning (Blaye & Light, 1995), Palincsar & Brown, 1984; Wertsch, 1991).

- *Mutual regulation... mutual diagnosis... we are back to the starting point!*

- Right, stop the recorder and send the tape to Jim.

Conclusion

Viewing cognition as distributed does not simply impact on the design of learning activities, but on the whole architecture of an ILE. Instead of viewing an ILE as a set of agents which exchange messages, the system and the learner are viewed as forming a joint cognitive system. The main functions of an ILE (tutoring, explaining, problem solving, diagnosis,...) are intrinsically collaborative. Cognitive effects are due to the mechanisms (e.g. appropriation) by which, despite the fact that cognition is distributed, agents succeed in forming a joint cognitive system. I illustrated this with the diagnosis process. While ILE researchers traditionally implemented diagnosis as a one-side process, dialogue models treat it as a mutual process. This does not simply imply that we should plug in a grounding dialogue module on top of the existing architecture. The growing importance of dialogue studies in AI&ED is not simply targeted to develop natural language interfaces. These techniques are required to model the key point in the 'distributed cognition' approach: participating into a joint cognitive system impacts on individual's cognition. I argue for a modification of the technology of knowledge-based systems in such a way that dialogues and problem solving are conducted by the same operators, that the very process of instantiation variable is implemented as a negotiation process.

Acknowledgements Thanks to Michael Baker, Patrick Mendelsohn and David Traum for their comments on previous drafts of this extended abstract, to the students who used MEMOLAB or PEOPLE POWER and to those who transcripted the videotapes. Parts of this work were supported by grants 4023-27013 and 1114-040711 from the Swiss National Science Foundation.

References.

Baker, M. (1992) The collaborative construction of explanations. Paper presented to "Deuxièmes Journées Explication du PRC-GDR-IA du CNRS", Sophia-Antipolis, June 17-19, 1992.

Baker, M. (1994) A model of negotiation in teaching-learning dialogues. *Journal of Artificial Intelligence in Education*, 5 (2), pp. 199-254.

Blaye, A. & Light, P. (1995). Collaborative problem solving with Hypercard: The influence of peer interaction on planning and information handling strategies. In C.E. O'Malley (Ed), *Computer-Supported Collaborative Learning. (pp.3-22).* Springer-Verlad: Berlin

Bull, S., Pain, H. & Brna, P. (1993) Collaboration and reflection in th econstruction of a student model for intelligent computer assisted language learning. In P*roceedings of the Seventh International PEG Conference*, Edinburgh.

Burton, R.T. (1982) Diagnosing bugs in a simple procedural skill. In Sleeman, D.H. and Brown J.S. (Eds) Intelligent Tutoring Systems. Academic Press, LOndon

Cawsey, A. (1991) A belief revision model for repair sequences in dialogue. In E. Costa (Ed), *New Directions for Intelligent Tutoring Systems*, pp. 183-197. Berlin: Springer-Verlag.

Cialdea, M. (1991) Meta-Reasoning and Student Modelling. In E. Costa (Ed), *New Directions for Intelligent Tutoring Systems*, pp. 71-90. Berlin: Springer-Verlag.

Clancey, W.J (1993) Guidon-Manage Revisited: A Socio-Technical Systems Approach. *Journal of Artificial Intelligence in Education*, vol. 4, n°1, pp. 5-30.

Clark, H.H., & Marshall, C.R. (1981) Definite reference and mutual understanding. in A.K. Joshi, B. L. Webber & I.A. Sag (Eds) *Elements of discourse understanding* (pp. 10-63). Cambridge: Cambridge University Press.

Clark, H.H., & Brennan S.E. (1991) Grounding in Communication. In L. Resnick, J. Levine & S. Teasley (Eds.), *Perspectives on Socially Shared Cognition* (127-149). Hyattsville, MD: American Psychological Association.

Dalal P.N. & Kasper, G.M. (1994) The design of joint cognitive systems: the effect of cognitive coupling on performance. International Journal of Human-Computer Studies, 40, pp. 677-702.

Dillenbourg, P. & Schneider D. (1993) Designers did not wait for situationists. A response to Clancey's Viewpoint "Guidon-Manage Revisited: A Socio-Technical Systems Approach". *Journal of Artificial Intelligence in Education*, vol. 4, n°1, pp. 41-48.

Dillenbourg, P., & Self, J.A. (1992) A computational approach to socially distributed cognition. European Journal of Psychology of Education, 3 (4), 353-372.

Dillenbourg, P., Mendelsohn P & Schneider D. (1994) The distribution of pedagogical roles in an intelligent learning environment. In B. Lewis and P. Mendelsohn (Eds) *Lessons From Learning* (pp. 199-216). Amsterdam: North-Holland (IFIP)

Dillenbourg, P. (to appear) Distributing cognition over brains and machines. In S. Vosniadou, E. De Corte, B. Glaser & H. Mandl (Eds), *International Perspectives on the Psychological Foundations of Technology-Based Learning Environments*. Hamburg: Springer-Verlag.

Dillenbourg, P., Baker, M., Blaye, A. & O'Malley, C. (to appear). The evolution of research on collaborative learning. In E. Spada & P. Reiman (Eds) *Learning in Humans and Machines*.

Douglas, S.A. (1991) Tutoring as Interaction: Detecting and Repairing Tutoring Failures. In P. Goodyear (Ed) *Teaching Knowledge and Intelligent Tutoring* (pp. 123-148). Norwood, NJ: Ablex.

Engel, F.L. and Haakma, R. (1993) Expectations and feedback in user-system communication. *International Journal of Man-Machine Studies*, 9, 427-452.

Fox, B.A. (1991) Cognitive and Interactional Aspects of Correction in Tutoring. In P. Goodyear (Ed) *Teaching Knowledge and Intelligent Tutoring* (pp. 149-172). Norwood, NJ: Ablex.

Gasser, L. (1991) Social conceptions of knowledge and action: DAI foundations and open systems semantics. *Artificial Intelligence*, 47, pp. 107-138.

Hirst, G, McRoy, S., Heeman, P. Edmonds, P. & Horton, S: (1994) Repairing Conversational Misunderstanding and Non-Understandings. *Speech Communication*, 12, 213-229.

Goldman, S.R., Petrosino, A., Sherwood, R.D., Garrison, S. Hickey, D.T., Bransford, J.D. and Pellegrino, J.W. (1994). Multimedia environments for enhancing science instruction. In S. Vosniadou, E. De Corte & H. Mandl (Eds),*Technology-Based Learning Environments*. PSychological and educational foundations. (pp. 89.96). Hamburg: Springer-Verlag

Newman, D. (1989) Is a student model necessary? Apprenticeship as a model for ITS. *Proceedings of the 4th AI & Education Conference* (pp.177-184), May 24-26. Amsterdam, The Netherlands: IOS.

Newman, D., Griffin P., & Cole M. (1989) The construction zone: working for cognitive change in school. Cambridge University Press: Cambridge.

Palincsar A.S. and Brown A.L. (1984) Reciprocal Teaching of Comprehension-Fostering and Comprehension-Monitoring Activities. *Cognition and Instruction,* vol.1, n°2, pp. 117-175.

Pea, R.D. (1993a) Learning scientific concepts throuhg material and social activities: conversational analysis meets conceptual change. Educational psychologist, 28 (3), pp. 265-277.

Pea, R.D. (1993b) Practices of distributed intelligence and designs for education. In G. Salomon (Ed) *Distributed Cognition*. Cambridge University Press, New York.

Resnick, L.B. (1991) Shared Cognition: Thinking as Social Practice. In L. Resnick, J. Levine and S. Teasley. *Perspectives on Socially Shared Cognition* (pp. 1-22). Hyattsville, MD: American Psychological Association.

Rogoff, B. (1991) Social interaction as apprenticeship in thinking: guided participation in spatial planning. In L. Resnick, J. Levine & S. Teasley (Eds). *Perspectives on Socially Shared Cognition* (pp. 349-364). Hyattsville, MD: American Psychological Association.

Roschelle, J. (1990) *Designing for Conversations*. Paper presented at the AAAI Syymposium on Knowledge-Based Environments for Learning and Teaching, Standford,CA, March 90.

Roschelle, J. & Teasley S.D. (1995) The construction of shared knowledge in collaborative problem solving. In C.E. O'Malley (Ed), *Computer-Supported Collaborative Learning. (pp. 69-197).* Berlin: Springer-Verlag

Self, J.A. (1988) Bypassing the intractable problem of student modelling. *Proceedings of the International Conference on Intelligent Tutoring Systems* (pp.18-24), June 1-3. Montreal, Canada.

Self, J.A. (1989) The case for formalising student models (and intelligent tutoring systems generally). Paper given at the *4th Conference on Artificial Intelligence and Education, Amsterdam*, The Netherlands.

Self, J.A: (1992) Computational Viewpoints. In R. Moyse & M. Elsom-Cook (Eds) *Knowledge Negotiation*.(pp. 21-40). London: Academic Press.

Resnick, L.B. (1991) Shared Cognition: Thinking as Social Practice.In L. Resnick, J. Levine and S. Teasley. *Perspectives on Socially Shared Cognition* (pp. 1-22). Hyattsville, MD: American Psychological Association.

Suchman L.A. (1987) Plans and Situated Actions. The problem of human-machine communication. Cambridge University Press. Cambridge.

Traum, D.R. (1994) A computational theory of grounding in natural language conversation. PhD Thesis & Technical Report 545. University of Rochester. Department of Computer Science.

Webb, N.M (1989) Peer interaction and learning in small groups. *International journal of Educational research*, 13 (1), 21-39.

Webb, N.M. (1991) Task related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education*, 22 (5), 366-389.

Wertsh, J.V. (1985) Adult-Child Interaction as a Source of Self-Regulation in Children. In S.R. Yussen (Ed).*The growth of reflection in Children* (pp. 69-97). Madison, Wisconsin: Academic Press.

Woods, D.D. & Roth E.M. (1988) Cognitive systems engineering. In M. Helander (Ed.) *Handbook of of Human-Computer Interaction* (pp. 3-43). Elsevier Science Publishers: Amsterdam.