DILLENBOURG P. (1996) 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.* (Pp. 165-184). Mahwah, NJ: Lawrence Erlbaum.

Distributing cognition over humans and machines

Pierre Dillenbourg [1]
TECFA, Faculty of Psychology and Educational Science,
University of Geneva
Switzerland

Abstract. This chapter considers computer-based learning environments from a socio-cultural perspective. It relates several concepts from this approach with design principles and techniques specific to learning environments. We propose a metaphor intended to help designers of learning environments to make sense of system features within the socio-cultural perspective. This metaphor considers the software and the learner as a single cognitive system, variably distributed over a human and a machine.

Keywords. Computer based learning, collaborative learning, distributed cognition, artificial intelligence, socio-cultural approach.

1. The Socio-Cultural Approach

The socio-cultural approach to human cognition has recently gained influence in the field of educational technology. This emergence can be explained by the renewed interest in America for Vygotsky's theories since the translation of his book (Vygotsky, 1978) and by the attacks against the individualistic view of cognition that dominated cognitive science (Lave, 1988). Moreover, the actual use of computers in classrooms leads scientists to pay more attention to social factors: teachers often have to put two or more students in front of each computer because schools generally have more students than computers! This situation was originally viewed as a restriction to the potential of computer-assisted instruction, since it was contradictory to the principle of individualization. Today, it is perceived as a promising way of using computers (Blaye, Light, Joiner and Sheldon, 1991).

The socio-cultural approach postulates that, when an individual participates in a social system, the culture of this social system and the tools used for communication, especially the language, shape the individual's cognition, and constitute a source of learning and development. The social influence on individual cognition can be analyzed at various levels: participation in a dyadic interaction (hereafter inter-psychological plane), participation in a 'community of practice' (e.g. colleagues) (Lave,

1 Address for correspondance: TECFA, PFSE, Université de Genève, Route de Drize, 9, 1227 Carouge. Switzerland. E-mail: pdillen@divsun.unige.ch

1991), and participation in increasingly larger social circles until the whole society and its inherited culture is included (Wertsch, 1991). In the dyadic interaction, one also discriminates between studies of collaboration between peers (i.e. subjects with even skills) and studies of apprenticeship (where one partner is much more skilled than the other). Within the socio-cultural perspective, one can examine interactive learning environments from different angles:

(1) The user-user interaction as a social process, mediated by the system.

When two human users (two learners or a learner and a coach) interact through the network or in front of the same terminal, the system influences their interaction. How should we design systems that facilitate human interaction and improve learning? Which system features could, for instance, help the co-learners to solve their conflicts? This viewpoint has been adopted in 'computer-supported collaborative learning'. It is receiving a great deal of attention because of the increasing market demand for 'groupware' (Norman, 1991).

(2) The user-designer relation as a social process, mediated by the system.

When a user interacts with a system (e.g. a spreadsheet), his reasoning is influenced by the tools available in this system. These tools embody the designer's culture. How should we design tools in such a way that users progressively 'think in terms of these tools' (Salomon, 1988) and thereby internalize the designers' culture? This viewpoint relates to the concept of *semiotic mediation* proposed by Wertsch (1991) to extend Vygotsky's framework beyond the inter-psychological plane.

(3) The user-system interaction as a social process.

When the learner interacts with a computerized agent performing (from the designer's viewpoint) a social role (a tutor, a coach, a co-learner,...), does this interaction have a potential for internalization similar to human-human conversations (Forrester, 1991; Salomon, 1990)? If the answer is yes, how should we design these agents to support learning ?

This chapter concentrates on the third viewpoint: the design of computerized agents which are engaged with the learner in a 'pseudo-social' relationship. One could object that the discrimination between the second and third view, i.e. the extent to which a program is considered as a tool (second view) or as an agent (third view) is purely metaphorical. Of course, it is. The 'tool' and 'agent' labels are images. Agents are supposed to take initiatives while tools are only reactive, but initiatives can be interpreted as sophisticated responses to previous learner behaviours. Actually, it is the user who determines whether he feels involved or not in a social relation with the machine: *"... the personification of a machine is reinforced by the way in which its inner workings are a mystery, and its behaviour at times surprises us"* (Suchman, 1987, p. 16). This issue is even more complex since many Intelligent Learning Environments (ILEs) include both tools and agents. For instance, People Power (Dillenbourg, 1992a) includes both a microworld and a computerized co-learner. However, the first experiments with this ILE seems to indicate that learners are able to discriminate when the machine plays one role or the other.

The main impact of the socio-cultural approach on ILEs is the concept of an 'apprenticeship' system. The AI literature refers to two kinds of apprenticeship systems: expert systems which apply machine learning techniques to integrate the user's solutions (Mitchell, Mabadevan and Stienberg, 1990) and learning environments in which it is the human user who is supposed to learn (Newman, 1989). We refer here to the latter. For Collins, Brown and Newman (1989), apprenticeship is the most widespread educational method outside school: in schools, skills are abstracted from their uses in the world, while in apprenticeship, skills are practised for the joint accomplishment of tasks, in their 'natural' context. They use the concept of 'cognitive apprenticeship' to emphasize two differences from traditional apprenticeship: (1) the goal of cognitive apprenticeship is to transmit cognitive and metacognitive skills (while apprenticeship has traditionally been more concerned with concrete objects and behaviors); (2) the goal of cognitive apprenticeship is that the learners progressively 'decontextualize' knowledge and hence become able to transfer it to other contexts.

2. The metaphor: Socially Distributed Cognition

This chapter is concerned with the relation between the socio-cultural approach and learning environments. We review several concepts belonging to the socio-cultural vocabulary and translate them in terms of the features found in learning environments. These concepts are considered one by one for the sake of exposition but they actually form a whole. To help the reader to unify the various connections we establish, we propose the following metaphor (hereafter referred to as the SDC metaphor):

View a human-computer pair (or any other pair) involved in shared problem solving as a single cognitive system

Since the boom in object-oriented languages, many designers think of the ILE as a multi-agent system. Similarly, some researchers think of a human subject as a society of agents (Minsky, 1987). The proposed metaphor unifies these two views and goes one step further. It suggests that two separate societies (the human and the machine), when they interact towards the joint accomplishment of some task, constitute a society of agents.

Two notions are implicit in this metaphor. First, a cognitive system is defined with respect to a particular task: it is an abstract entity that encloses the cognitive processes to be activated for solving this particular task. The same task may be solved by several cognitive systems, but the composition of a cognitive system is independent of the number of people who solve the task. The second implicit notion is that agents (or processes) can be considered independently from their implementation (i.e. their location in a human or a machine): a process that is performed by a subject at the beginning of a session can be performed later on by his partner. Studies of collaborative problem solving have shown that peers spontaneously distribute roles and that this role distribution changes frequently (Miyake, 1986; 0'Malley, 1987; Blaye et al., 1991). We use the term 'device' to refer indifferently to the person or the system that performs some process.

The following sections attempt to clarify how this model relates to the socio-cultural framework at one end, and at the other end, what it means in terms of implementation.

3. Learning environments

In the remainder of this chapter, we will refer frequently to three systems we have designed: PEOPLE POWER (Dillenbourg, 1992a; Dillenbourg and Self, 1992), MEMOLAB (Mendelsohn, this volume) and ETOILE (Dillenbourg, Hilario, Mendelsohn, Schneider and Borcic, 1993). We briefly describe these systems now in order to make later references shorter. Some features of these systems make sense within the socio-cultural perspective, even though these systems were not designed specifically to address socio-cultural issues.

3.1. People Power

PEOPLE POWER is a learning environment in which the human learner interacts with an artificial learning companion, hereafter referred to as the 'co-learner'. Its pedagogical goal is that the human learner discovers the mechanisms by which an electoral system is more or less proportional. The system includes four components (see figure 1): (1) a microworld in which the learner can design an electoral experiment (i.e. choose parties, candidates, laws, etc.), run the elections and analyze the results; (2) an interface by which the human learner (and conceptually the co-learner) plays with the microworld; (3) the co-learner, named Jerry Mander, and (4) an interface that allows the human and the computerized learners to communicate with each other.

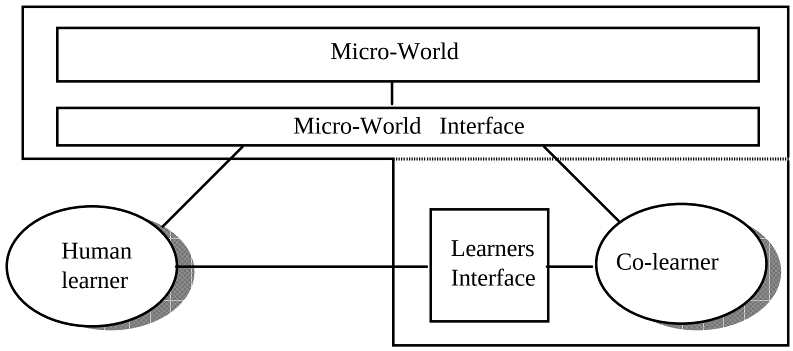

Figure 1: Components of human-computer collaborative learning system.

The learners play a game in which the goal is to gain seats for one's own party. Both learners play for the same party. They engage in a dialogue to agree on a geographical organization of wards into constituencies. The co-learner has some naive knowledge to reason about elections. This knowledge is a set of rules (or arguments). For instance, a rule says "If a party gets more votes, then it will get more seats". This rule is naive but not basically wrong, it is only true in some circumstances. The co-learner learns how to apply this rule when it reasons about the way to gain seats for its party.

We tested PEOPLE POWER under two paradigms: with two artificial learners and with a human learner and an artificial learner (Dillenbourg and Self, 1992). The human/artificial experiments were informal and restricted to five subjects. Learners appreciated the possibility of interacting with the co-learner. Some expressed the feeling of actually collaborating with a partner, though this partner did not exhibit completely human behaviour. We later report some observations that are related to the SDC metaphor.

3.2. MEMOLAB / ETOILE

The goal of MEMOLAB is for psychology students to acquire the basic skills in the methodology of experimentation. The learner builds an experiment on human memory. A typical experiment involves two groups of subjects each encoding a list of words. The two lists are different and these differences have an impact on the recall performance. An experiment is described by assembling events on a workbench. Then, the system simulates the experiment (by applying case-based reasoning techniques on data found in the literature). The learner can visualize the simulation results and perform an analysis of variance.

This artificial lab constitutes an instance of a microworld. Most learners need some external guidance to benefit from such a microworld. We added computational agents (coach, tutors and experts) to provide this guidance. But we also explored another way of helping the learner: by structuring the world. MEMOLAB is actually a sequence of microworlds. The relationship between the objects and operators of two successive microworlds parallels the relationship between developmental stages in the neo-piagetian theory of Case (Mendelsohn, this volume). At the computational level, the relationship between successive worlds is encompassed in the interface: the language used in a microworld to describe the learner's work is used as the command language for the next microworld. This relationship, referred to as a 'language shift', will be explained in section 4.5.

A goal of this research project was to generalize the solutions developed for MEMOLAB and to come out with a toolbox for creating ILEs. We achieved domain independence by defining teaching styles as a set of rules which activate and *monitor the interaction between an expert and the learner*. The technical solutions chosen for obtaining a fine-grained interaction between the expert and the learner will be described in section 4.2. This toolbox is called ETOILE (Experimental Toolbox for Interactive Learning Environments).

4. From concepts to systems

In this section, we review several key concepts from the socio-cultural approach and attempt to translate them in terms of ILE design. We therefore use the proposed metaphor: view two interactive problem solvers as a single society of agents.

4.1. Zone of proximal development, scaffolding and fading

We start our review of socio-cultural concepts with Vygotsky's (1978) concept of '*zone of proximal development*' (ZPD). The ZPD is the difference between the child's capacity to solve a problem alone and his ability to solve it under adult guidance or in collaboration with a more capable peer. Although it was originally proposed for the assessment of intelligence, it nowadays inspires a great deal of instructional organisation (Wertsch, 1991). *Scaffolding* is the process of providing the learner with the help and guidance necessary to solve problems that are just beyond what he could manage independently (i.e. within his ZPD). The level of support should progressively decrease (fading) until the learner is able to solve the problem alone.

The process of scaffolding has been studied by Rogoff (1990, 1991) through various experiments in which children solved a spatial planning task with adults. She measured the performance of children in a post-test performed without adult help. She established a relationship between the type of adult-child interactions and the post-test results. Children scored better in the post-test in the cases where the problem solving strategy was made explicit by the adult. These results are slightly biased by the fact that the proposed task (planning) is typically a task in which metaknowledge plays the central role. Nevertheless, on the same task, Rogoff observed that children who worked with an adult performed better than those who worked with a more skilled peer. Similarly, she found that efficient adults involved the child in an explicit decision process, while skilled peers tended to dominate decision making.

In terms of the SDC metaphor, scaffolding can be translated as activating agents that the learner does not or cannot activate. Fading is interpreted as a quantitative variation of the distribution of resources: the number of agents activated by the machine decreases and the number of agents activated by the learner increases. In ETOILE for instance, a teaching style determines the quantitative distribution of steps among the expert and the learner (and its evolution over time). However, Rogoff's experiments show it is not relevant to count the number of agents activated by each partner, unless we take into consideration the hierarchical links between agents. Some agents are more important than others because they play a strategic role: when solving equations, the agent 'isolate X' will trigger several subordinated agents such as 'divide Y by X'. This implies that the agents society must be structured in several control layers. The issue of making control explicit has been a key issue for several years in the field of ILEs (Clancey, 1987). In other words, fading and scaffolding describe a variation in learner control, but this variation does not concern a quantitative ratio of agents activated by each participant. It refers to the qualitative relationship between the agents activated on each side.

Tuning the machine contribution to the joint accomplishment of a task may affect the learner 's interest in collaboration. What one can expect from a partner partially determines one's motivation to collaborate with him. The experiments conducted with People Power showed interesting phenomena of this kind. Initially, the subjects who collaborated with the machine did not always accept that the computer was ignorant. Two subjects even interrupted their session to tell us that the program was buggy. They were surprised to see a computer suggesting something silly (though we announced that this would be the case). Later on, subjects appeared to lose their motivation to collaborate if the co-learner was not improving its suggestions quickly enough. Our machine-machine experiments showed that the co-learner performance depended on the amount of interactions among learners. In People Power, the cost of interaction with the co-learner was very high. The subjects reduced the number of interactions and hence the co-learner learned slowly. All dialogue patterns elaborated by the co-learner during these one-hour sessions were much more rudimentary that the patterns built with another artificial learner (where there was no communication bottle-neck). These patterns depend on the quantity and variety of interactions. They determine the level of elaboration of Jerry's arguments and hence the correctness of its suggestions. Jerry continued to provide the learners with suggestions that were not very good, and decreased the subjects' interest in its suggestions. In terms of the SDC model, these observations imply that *the agents implemented on the computer should guarantee some minimal level of competence for the whole distributed system, at any stage of scaffolding/fading.* This 'minimal' level is the level below which the learner loses his interest in interacting with the machine.

4.2. Participation and appropriation

A core idea in the socio-cultural approach is the notion of participation: "the skills a student will acquire in an instructional interaction are those required by the student's role in the joint cognitive process." (Bereiter and Scardamalia, 1989, p. 383). The challenge is to understand why participation in joint problem solving may sometimes change the understanding of a problem. Rogoff (1991) explains it by a process of 'appropriation'. Appropriation is the socially-oriented version of Piaget's biologically-originated concept of assimilation (Newman, Griffin and Cole, 1989). Appropriation is mutual: each partner gives meaning to the other's actions according to his own conceptual framework. Appropriation constitutes a form of feed-back: if two persons, A and B, interact, when A performs the first action and B the next one, B's action indicates to A how his first action is interpreted by B. In other words, B's action is information on how A's action makes sense within B's conceptualization of the problem. Fox (1987) reported that humans modify the meaning of their action retrospectively, according to the actions of others that follow it. This form of feed-back requires that problem solvers are opportunistic, i.e. able to escape from an established plan in order to integrate their partner's contribution into their own solution path.

The difference between this kind of feed-back and the behaviourist kind of feed-back is that the partner may have no didactic intention. In MEMOLAB, the expert's actions are purely egocentric, the expert wants to solve the task. These actions may constitute some kind of feed-back for the learner, but the expert does not teach. This different conception of feed-back gives more importance to collaboration than to diagnosis

(Newman, 1989). We can illustrate the difference between diagnosis-based feedback and collaboration-oriented feed-back by the difference between a mal-rule and a repair rule.

- A mal-rule is something like 'if the problem state has these features, then do something wrong' or, more concretely, 'If you want to go from Paris to Brussels, fly West'. A mal-rule sets a hypothesis concerning the cognitive cause of an error. This concept has been frequently used in student modelling.

- In MEMOLAB, we used the concept of a repair rule to support certain types of expert-learner interactions. The format of a repair rule is 'if the problem state is wrong then correct it this way', or more concretely 'If you fly from Paris to Brussels and see London, then turn to the East'. As any ordinary rule, a repair rule generates some expert's action. Note that we use this term independently from Brown's and Van Lehn's use of it (Brown and Van Lehn, 1980).

The specificity of an interactive expert system is that interaction may lead the expert outside its normal solution path. A genuine expert rule-base generates an optimal solution path. If this expert collaborates with a human learner, it may encounter problem states that do not belong to this optimal path, but which are still correct. The expert therefore has sub-optimal rules that are considered after the optimal rules have been evaluated (rules have a priority parameter). Repair rules are concerned with a third situation, when the interaction leads the expert to an incorrect problem state that would belong neither to the optimal nor to the sub-optimal solution path.

One must be careful with regard to the benefits expected from interactions which look collaborative but where partners execute processes independently from each other. "The presence of a partner may be irrelevant, unless the partners truly share their thinking processes in problem solving." (Rogoff, 1991, p. 361). The SDC metaphor provides us with a framework within which we implemented a 'shared thinking process', i.e. we distributed sub-processes over partners. Sharing processes implies a high level of *granularity and opportunism*. Granularity refers to the size of agents. The limit for granularity is defined by a pedagogical criterion: learners must be able to talk about what an agent does. Opportunism means that each agent integrates in his reasoning any new element brought by its partner.

How to obtain a high level of granularity and opportunism? Technical solutions are emerging in 'distributed artificial intelligence' (Durfee, Lesser and Corkill, 1989) for collaboration between computational agents. Regarding collaboration between a human and a machine, we learned from our observations made with subjects using PEOPLE POWER. As we designed it, the object of discussion was the set of arguments that constituted Jerry's naive knowledge. Actually, we observed that the main activity of learners was to look at the table showing the data for each party in each ward. Reasoning about the effect of moving a ward was more than simply manipulating arguments or rules, it was mentally moving a column of a table and re-computing the sums to check whether this new ordering led to the gain of a seat. The interaction with the co-learner would have been more relevant if it had been about moving columns. The interface should, for instance, provide facilities to move columns and recompute sums by constituencies. Hence, in MEMOLAB, we have chosen a solution based on what can be the most easily shared between a person and a machine: the interface. Let us imagine two production systems that use a common set of facts. They share the same representation of the problem. Any fact produced by one of them is added to this common set of facts. Hence, at the next cycle of the inference engine, this new fact may trigger a rule in either of the two rule-bases. Now, let us replace one computerized expert by a human learner. The principle may still apply provided we use an external problem representation instead of the internal one. The common set of facts is the problem representation as displayed on the interface (see figure 2).

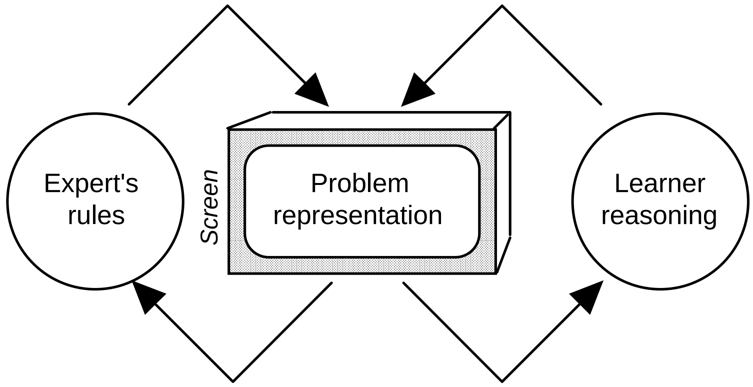

Figure 2: Opportunism in human-machine collaboration in MEMOLAB

All the conditions of the machine's rules refer only to objects displayed on the screen.[2] The actions performed by the rules modify the problem representation. In short, the shared representation is visible by both partners and can be modified by both partners. We do not claim that they share the same 'internal' representation. Sharing an external representation does not imply at all that both partners build the same internal representation. The shared concrete representation simply facilitates the discussion of the differences between the internal representations and hence improves the co-ordination of actions.

4.3. Internalization

The mechanism underlying the appropriation process is the process of internalization, a central concept in Vygotsky's framework: "Every function in the child's development appears twice: first, on the social level, and later on the individual level; first, between people (inter-psychological) and then inside the child (intra-psychological)" (Vygotsky, 1978). Internalization refers to the genetic link between the social (or inter-psychological) and the inner (or intra-psychological) planes. Social speech is used for interacting with humans, inner speech is used to talk to ourselves, to reflect, to think. Inner speech has a function of self-regulation.

The SDC model is based on the relationship between social speech and inner speech: if an individual and a group are both modelled as a society of agents, inner speech and social speech are two instances of communication among agents. Inner speech is communication among agents implemented on the same device (intra-device communication), social speech occurs between agents belonging to different devices (inter-device communication). These levels of speech are two instances of the class 'communication among agents'. There is however a difference: communication between agents from different devices is external, it is observable. Patterns of inter-device communication can hence be induced and applied to intra-device communication. These ideas can be summarized as follow:

1. An individual is a society of agents that communicate. A pair is also a society, variably partitioned into devices.

---

2 Technically, we use an object-oriented inference engine. Any rule variable is defined with respect to a class. This variable can only be instantiated by the instances of its attached class. Objects of class X that are displayed on the screen are actually defined as instances of a subclass of X, the class Displayed-X.

2. The device border determines two levels of communication: agent-agent communication and device-device communication. Inter-agent and inter-device communications are isomorphic.

3. Inter-device communication is observable by each device. Therefore, inter-device communication patterns generate intra-device communication patterns.

The three postulates of the SDC model have been implemented for designing JERRY MANDER, the computerized co-learner in PEOPLE POWER. Jerry talks with the human learner (or another artificial learner) about the relation between the features of an electoral system and the results of elections. During this dialogue, Jerry stores relations between arguments. A network of arguments constitutes what we called a 'communication pattern'. Let us assume Jerry claims that Ward-5 should be moved from 'Southshire' to 'Northshire', because this would increase the score of his party in 'Northshire'. His partner may raise the objection that the party would lose votes in 'Southshire'. Jerry will then create a 'refutation' link between the first argument and its counter-argument. Similarly, dialogue patterns include 'continue-links' between arguments that have been verbalized consecutively in a successful argumentation. Jerry Mander reuses these dialogue patterns when it reasons alone. For instance, when it considers another move, Jerry Mander retrieves the counter-argument connected by a refutation-link and checks whether this counter-argument is valid in the new context. If it is, it refutes itself[3]. Using a refutation-link between arguments (stored as rules) corresponds to a mechanism of specialization (adding a condition), while the use of continue-links corresponds to a form of chunking.

To implement the isomorphism between inner speech and social speech (second axiom of the SDC model), we used the following trick: Jerry Mander uses the same procedure for talking with his partner and talking with itself. It uses a single procedure 'dialogue' which accepts two entries, a 'proposer' and a 'refuter'. The procedure call 'dialogue learner-X learner-Y' gives a real dialogue while the call 'dialogue learner-X learner-X' corresponds to individual reasoning (monologue). The implementation is actually a bit more complex. Each link is associated with a description of the context in which the connected arguments have been verbalized and has a numeric weight. This numeric weight evolves according to the partner's agreement ('social sensitivity') and to the electoral results ('environmental sensitivity'). A complete description of the learning mechanisms and their performance can be found in Dillenbourg and Self (1992).

4.4. Private speech and reification

Between inner speech and social speech, psychologists discriminate an intermediate level termed 'egocentric speech' by Piaget and 'private speech' by Vygotsky. These concepts are not completely synonymous (Zivin, 1979). The most familiar examples of private or egocentric speech are the conversations conducted aloud by children who play alone. We might also refer to the verbal productions of people using computers on their own. Egocentric or private speech still has a rather social form (it is verbalized, it has some syntax, ...), but it has lost its social function (it is produced in the absence of any other person). For Piaget, it corresponds to some kind of uncontrolled production, while for Vygotsky, it has a self-regulating function (Zivin, 1979). The interest in this intermediate level is that psychologists may extrapolate differences between social and private speech to speculate on the features of inner speech.

3 It backtracks in the process of proving that its proposition will lead to gaining a seat.

There is an interesting similarity between private speech and the idea of *reification*, a technique used in ILEs to support reflection. Reflection, i.e. the process of becoming aware of one's own knowledge and reasoning, is receiving growing attention from ILE designers (Collins and Brown, 1988). Systems that attempt to promote reflection often present some trace of the learner's activities and of the environment's responses. Systems such as ALGEBRALAND (Collins and Brown, 1988) or the GEOMETRY TUTOR (Anderson, Boyle and Yost, 1985) facilitate the learner's reflection by displaying the learner's solution path as a tree structure. This representation shows that solving an equation or proving a theorem are not straightforward processes, but require numerous attempts and frequent backtracking. Such a representation reifies, makes concrete, some abstract features of the learner's cognition. It is not neutral, but results from an interpretation by the machine of the learner's action. Through this interpretation, the learner can understand how his action makes sense within the system conception of the domain (see section 4.2 on appropriation).

This graphical representation of behaviour has the ambiguous status of private speech. In systems such as ALGEBRALAND, TAPS (Derry, 1990) or HERON (Reusser Kampfer, Sprenger, Staub, Stebler and Stussi, 1990), this graphical representation serves both as a way of communicating with the system and as a tool for self-regulation. One can argue whether Vygotsky's theories on (verbal) speech are compatible with graphical languages used in modern interfaces. For Salomon (1988), "all tools that we consider as prototypical and as intuitively appealing candidates for internalization have also a distinctive spatial form".

4.5. Language shift

Wertsch (1985) reports an interesting study which investigates the role of languages in internalization. This study zooms in the inter-psychological plane, observing mothers helping their children (2 1/2 and 3 1/2 year old) to construct a puzzle in accordance with a model (the same puzzle already built). Wertsch contrasted the language used by mothers to refer to puzzle pieces according to the child's age: mothers of younger children designate directly the piece by pointing to it or by its colour (e.g. "a green piece"), while mothers of older children refer to pieces with respect to the problem-solving strategy (e.g. "one colour that's not the same"). For Wertsch, the cognitive processes required to participate in a strategy-oriented dialogue are virtually equivalent to the cognitive processes necessary to apply this strategy without the mother. This study confirms Rogoff's view that participation changes understanding. It weakens the dichotomous distinction between the social and internal planes, since changes inside the social plane may be more important than the social-internal transition.

We suggested that a mechanism of shifting between two language levels, as observed by Wertsch, could be applied to the design of ILEs (Dillenbourg, 1992b). Let us decompose the difference between a novice and an expert into several levels of skills. When learners solve problems at level X, they interact with the system through some command language CL_x. The system, as the mothers in Wertsch's observations, has a second language available, called the description language (DL). The description language reifies the implicit problem solving strategy by displaying a trace of the learner's activities. The system uses this description language in feed-back in order to associate the two descriptions of the same solution, one expressed in the command language and the second in the description language. The language shift occurs when the system moves up in the hierarchy of expertise levels. After a move to the next level (X+1), learners receive a new command language which includes the concepts that were introduced by the description language at level X. This language shift can be expressed by the equation $CL_{x+1} = DL_x$. After the language shift, learners are compelled to use explicitly the operators or strategies that where previously implicit (but reified in the description language). We illustrate this principle with two instances:

- Let us imagine a courseware for learning to solve equations. At the first level, the learner would manually perform algebraic manipulations (CL1). Sequences of transformations would be redisplayed as higher order operators, such as 'move X to LHS' (DL1). After the language shift, the learner would transform equations by applying directly these operators (CL2 = DL1).

- This principle has been applied to the definition of the microworlds in MEMOLAB (Mendelsohn, this volume). MEMOLAB includes three successive microworlds or levels, with increasingly powerful command languages (and increasingly complex problems to solve). At level 1, a psychological experiment is built by assembling chronological sequences of concrete events. At level 2, the learner does not describe each concrete event but builds the treatment to be applied to each group of subjects. At level 3, the learner directly defines the experimental plan, i.e. the logical structure of the experiment. At levels 1 and 2, when an experiment design has been completed, the experiment is 'redisplayed' with the formalism used at the next level.

4.6. Social grounding

In PEOPLE POWER, the internalization mechanism has been implemented as the simple storage of dialogue patterns. It is clear however that internalization is not a simple recording process, but a transformation process. There exist interesting similarities between the transformations which occur during internalization and those which occur during social grounding, although these processes been studied independently from each other. Social grounding is the mechanism by which two participants in a discussion try to elaborate the mutual belief that their partner has understood what they meant to a criterion sufficient for the current purpose (Clark and Brennan, 1991).

For Vygotsky, inner speech has a functional similarity with social speech but loses its structural similarity. For Luria (1969), "inner speech necessarily ceases to be detailed and grammatical" (p. 143). The difference between social and inner speech is due to the fact that "inner speech is just the ultimate point in the continuum of communicative conditions judged by the degree of 'intimacy' between the addresser and addressee" (Kozulin, 1990, p. 178). The main transformation observed as a result of the intimacy between the addresser and addressee is a process of *abbreviation* (Kozulin, 1990; Wertsch, 1979, 1991). Interestingly, Krauss and Fussell (1991) found the same abbreviation phenomena in social grounding. They report several experiments in which subjects have to establish expressions to refer to 'non sense figures' and to use these references later on, in various conditions. Subjects first refer to a particular picture by saying "Looks like a Martini glass with legs on each side". Later on, this reference becomes "Martini glass with the legs", then "Martini glass shaped thing", then "Martini glass" and finally "Martini". Krauss and Fussell show that the decrease in expression length is a function of the feed-back given by the listener. Another interesting experiment compares expressions built by the subjects for themselves or for peers. They observe that personal messages were less than half as long as social messages. Why does abbreviation occur both during internalization and during social grounding? The explanation may be that internalization and social grounding are two phenomena during which the addresser acquires information about the addressee's understanding of his messages.

The work on grounding is very important for designers of ILEs, it is even at the heart of recent debates in artificial intelligence. The symbol grounding crisis (Harnad, 1990) launched intensive research to design situated robots (Maes, 1990), i.e. robots which can physically ground their symbols in the environment, through actors and sensors. Less attention has been paid to the possibility of social grounding, i.e. grounding the system symbols in the user's experience. Any communication is ambiguous, it works because humans constantly detect and repair communication breakdowns, but computers have far fewer resources than humans for detecting and repairing communication failures (Suchman, 1987). Previous sections placed great expectations on the cognitive benefits of using graphical languages, but how do we guarantee that learners correctly interpret these symbols? Wenger (1987) defined the *epistemic fidelity* of a representation as the degree of consistency between the physical representation of some phenomena and the expert's mental representation of this phenomena. Roschelle (1990) attempted to apply this principle to the design of the Envisioning Machine (EM), a direct-manipulation graphical simulation of the concepts of velocity and acceleration. He successively designed several representations for the same set of physical phenomena (particle movements). The first EM design focused on epistemic fidelity. However, because mapping physical and mental representations is an inherently ambiguous interpretation process, the users did not read representations as experts did. Representations do not hold some trivial meaning but, inversely, can be used to support social grounding. Roschelle (1992) observed that learners use the computer to test under increasingly tighter constraints the degree to which their interpretation of physical phenomena were shared. Roschelle refers to this property of representations as 'symbolic mediation'.

Two implications can be derived. The first has been quoted before: collaboration should be concerned with what is happening on the screen (not some hidden knowledge), since the screen is the main reference to establish shared meanings (see section 4.2). The second implication is that dialogue should be more about understanding than about agreement. In PEOPLE POWER, our dialogue patterns were rudimentary, including only agreement or disagreement. This indicates a Piagetian bias. Doise and Mugny (1984) have extended the notion of conflict between a subject's beliefs and the world events to include conflict between opposite beliefs held by different subjects. This socio-cognitive conflict is more likely to be perceived and solved because of the social pressure to maintain partnership. However, in her thesis, Blaye (1988) found very little evidence of real conflict in pairs. The concept of conflict is not operationally defined. Where is the frontier between a divergence of focus (Miyake, 1986), some disagreement (Blaye, 1988) and an open conflict? There is no clear answer to that question. Actually, disagreement in itself seems to be less important than the fact that it generates communication between peer members (Blaye, 1988; Gilly, 1989). Bearison et al. (1986, quoted by Blaye, 1988) reported that non-verbal disagreement (manifested for instance by moving the object positioned by the partner) was not predictive of post-test score increase. Blaye (1988) suggests that "oppositions give rise to verbalizations that regulate the partner's activities and may possibly contribute to the internalization, by the producer, of an adequate regulation mode."[4] (p. 398). The notion of conflict results from a bipolarisation of the continuum which goes from a total absence of understanding to fully shared understanding. Collaboration among agents should not be simplified to agreement or disagreement. It should be considered a complex social grounding process. Of course, artificial intelligence handles dialogue moves such as "continue" or "refute" more easily than it encodes meanings. Intermediate solutions could be to extend richer sets of dialogue moves able to generate sub-dialogues aiming to elaborate, disambiguate or repair communication (Baker, 1992). The person-machine interface must have some 'noise tolerance', some space where the meaning of concepts can be negotiated.

Some recent experiments with MEMOLAB (Dillenbourg et al., 1993) revealed mechanisms of human-machine grounding: the learner perceives how the machine understands him (i.e. he makes a diagnosis of the machine diagnosis) and reacts in order to correct eventual misdiagnosis. However, in the current implementation of

---

4 My translation.

MEMOLAB, rule variables unambiguously refer to screen objects. To support social grounding, the instantiation of rule variables by displayed objects should not be a completely internal process, but the result of some interaction with the learner.

5. Synthesis

Let us integrate these various connections between theories and systems within the SDC model. The learner and the system form a single cognitive system. This system is a society of agents, characterized by high granularity (agents have narrow skills). All agents are implemented on the machine and on the learner. Implementing a computerized agent-X 'on the learner' means designing an interface function by which the learner may perform the same operation as agent-X. The total number of activated agents remains constant, in such a way that, from the beginning, the learner and the computer jointly achieve meaningful and motivating tasks. The machine agents are activated to complement the agents activated by the learner (scaffolding). An agent activated by the learner is deactivated by the machine (fading). Some agents encompass and make explicit the problem solving strategy.

The agents interact about what is on the screen. The behaviour of the computerized agents is determined by what is on the screen. The interface is the permanent updated representation of the problem which serves as the basis for activating computerized agents. Agents use the interface as a reference to establish mutual understanding (social grounding). The communication among agents is not didactic in itself, it serves the accomplishment of the task, but it may indirectly fulfil a didactic function (appropriation).

The forms of communication among agents are inspired by the reasoning processes the learner should acquire: the system-learner interactions are the source of the learner's future self-regulation mechanisms. To support this internalization, reified graphics serve both for reflection and communication with machine agents (private speech).

More generally, we argue that metaphors are useful for designers. Designers often complain about the lack of formal models of learning and teaching (with a few notable exceptions), which would generate clear design specifications. We do not believe that the design of a learning environment will ever be a deductive process. Design is a creative process during which one attempts to make sense of learning activities or system features within a theoretical framework.

Acknowledgements

Thanks to P. Brna, S. Bull, M. Lee and to the editors for their comments of this chapter.

References

Anderson, J.R., Boyle, C.F., & Yost, G. (1985) The Geometry Tutor. Proceedings of the Ninth International Joint Conference on Artificial Intelligence (Vol. I, 1-7). Los Angeles. August 18-23.

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.

Bereiter, C., & Scardamalia, M. (1989) Intentional learning as a goal of instruction. In L.B. Resnick (Ed.), *Cognition and Instruction: Issues and Agendas* (361-392). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Blaye, A. (1988) *Confrontation socio-cognitive et resolution de problèmes.* Doctoral dissertation, Centre de Recherche en Psychologie Cognitive, Université de Provence, 13261 Aix-en-Provence, France.

Blaye, A., Light, P., Joiner, R., & Sheldon, S. (1991) Collaboration as a facilitator of planning and problem solving on a computer based task. *British Journal of Psychology*, 9, 471-483.

Brown, J.S., & Van Lehn, K. (1980) Repair theory: a generative theory of "bugs" in procedural skills. *Cognitive Science*, 4, 379-426.

Case, R. (1985) *Intellectual Developpement: from Birth to Adulthood*. New York: Academic Press.

Clancey, W.J. (1987) *Knowledge-based tutoring: the Guidon Program*. Cambridge, Massachusetts: MIT 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.

Collins, A., & Brown, J.S. (1988) The Computer as a Tool for Learning through Reflection. In H. Mandl & A. Lesgold (Eds), *Learning Issues for Intelligent Tutoring Systems* (1-18). New York: Springer Verlag.

Collins, A., Brown J.S., & Newman, S. (1989) Cognitive apprenticeship: teaching the craft of reading, writing and mathematics. In L.B. Resnick (Ed.), *Cognition and Instruction: Issues and Agendas* (453-494). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Derry, S.J. (1990) *Flexible Cognitive Tools for Problem Solving Instruction*. Paper presented at the AERA symposium, Computers as Cognitive Tools, April. Boston, MA.

Dillenbourg, P. (1992a) *Human-Computer Collaborative Learning*. Doctoral dissertation. Department of Computing. University of Lancaster, Lancaster LA14YR, UK.

Dillenbourg, P. (1992b) The Language Shift: a mechanism for triggering metacognitive activities. In P. Winne & M.Jones. *Adaptive Learning Environments.*: *foundations and frontiers*. Springer-Verlag. Hamburg

Dillenbourg, P., Hilario, M., Mendelsohn, P., Schneider, D., & Borcic, B. (1993) The Memolab Project. Research Report. TECFA Document. TECFA, University of Geneva.

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

Doise, W., & Mugny, G. (1984) *The social development of the intellect*. Oxford: Pergamon Press.

Durfee, E.H., Lesser, V.R., & Corkill, D.D. (1989) Cooperative Distributed Problem Solving. In A. Barr, P.R. Cohen & E.A. Feigenbaum (Eds.),*The Handbook of Artificial Intelligence*, (Vol. IV, 83-127). Reading, Massachusetts: Addison-Wesley.

Forrester, M.A. (1991) A conceptual framework for investigating learning in conversations. *Computers in Education*, 17 (1), 61-72.

Fox, B. (1987) Interactional reconstruction in real-time language processing. *Cognitive Science*, 11 (3), 365-387.

Gilly, M. (1989) The psychosocial mechanisms of cognitive constructions, experimental research and teaching perspectives. *International Journal of Educational Research*, 13, 6, 607 - 621.

Kozulin, A. (1990) *Vygotsky's psychology.* A biography of ideas. Harvester, Hertfordshire.

Krauss, R.M., & Fussell, S.R. (1991) Constructing shared communicative environments. In L. Resnick, J. Levine & S. Teasley (Eds.), *Perspectives on Socially Shared Cognition* (172-202). Hyattsville, MD: American Psychological Association.

Lave J. (1988) Cognition in Practice. Cambridge: Cambridge University Press

Lave J. (1991) Situating learning in communities of practice. In L. Resnick, J. Levine & S. Teasley (Eds.), *Perspectives on Socially Shared Cognition* (63 - 84). Hyattsville, MD: American Psychological Association.

Luria, A.R. (1969) Speech development and the formation of social processes. In M. Cole & I. Maltzman (Eds.), *A handbook of contemporary Soviet psychology.* New York: Basic Books.

Maes, P. (1990) Situated agents can have goals. *Robotics and Autonomous Systems*, 6, 49-70.

Minsky, M (1987) *The society of mind*. London: William Heinemann Ltd.

Mitchell, T.M., Mabadevan, S.M., & Stienberg, L.I. (1990) LEAP: A learning apprentice for VLSI design. In Y. Kodratoff & R.S. Michalski (Eds). *Machine Learning*. (Vol. III, 271-301). Palo Alto, CA: Morgan Kaufmann.

Miyake, N. (1986) Constructive Interaction and the Iterative Process of Understanding. *Cognitive Science*, 10, 151-177.

Newman, D. (1989) Is a student model necessary? Apprenticeship as a model for ITS. *Proceedings of the 4th AI & Education Conference* (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.

Norman, D.A. (1991) Collaborative computing: collaboration first, computing second. *Communications of the ACM*, Vol. 34 (12), 88-90.

O'Malley, C. (1987) *Understanding explanation*. Paper presented at the third CeRCLe Workshop, Ullswater, UK.

Piaget, J. (1928) *The language and thought of the child*. New York: Harcourt.

Reusser, K., Kampfer, A., Sprenger, M., Staub, F., Stebler, R., & Stussi, R. (1990) *Tutoring mathematical word problems using solution trees*. Research Report No 8, Abteilung Pädagogishe Psychologie, Universität Bern, Switzerland.

Rogoff, B. (1990) *Apprenticeship in thinking*. New York: Oxford University Press

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* (349-364). Hyattsville, MD: American Psychological Association.

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

Roschelle, J. (1992) Learning by Collaborating: Convergent Conceptual Change. *Journal of the Learning Sciences*, 2, 235-276.

Salomon, G. (1988) AI in reverse: computer tools that turn cognitive. *Journal of educational computing research*, 4, 12-140.

Salomon, G. (1990) Cognitive effects with and of computer technology. *Communication research*, 17 (1), 26-44.

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

Vygotsky, L.S. (1978) Mind in Society. *The Development of Higher Psychological Processes*. Edited by M. Cole, V. John-Steiner, S. Scribner & E. Souberman. Cambridge, Massachusetts: Harvard University Press.

Wenger, E. (1987) *Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge*. Los Altos, CA: Morgan Kaufmann

Wertsch, J. V. (1979) The regulation of human action and the given-new organization of private speech. In G. Zivin (Ed.), *The development of self-regulation through private speech*, 79-98. New York: John Wiley & Sons.

Wertsch, 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* (69-97). Madison, Wisconsin: Academic Press.

Wertsch, J.V. (1991) A socio-cultural approach to socially shared cognition. In L. Resnick, J. Levine & S. Teasley. *Perspectives on Socially Shared Cognition* (85 - 100). Hyattsville, MD: American Psychological Association.

Zivin, G. (1979) Removing common confusions about egocentric speech, private speech and self-regulation. In G. Zivin (Ed.), *The development of self-regulation through private speech*, 13-50. New York: John Wiley & Sons