15

Representations in practices

A sociocultural approach to multimodality in reasoning

Jonas Ivarsson, Jonas Linderoth and Roger Säljö

Introduction

In human history, representations play a significant role for the development of knowledge, skills and identities and for all the various kinds of social practices that go into community building and social life in general. As Luria (1981) and many others have pointed out, people do not just live *in* the world, we can also communicate *about* it (cf. Cole, 1996, p. 120). A necessary prerequisite for being able to communicate about the world is that we have access to symbolic representations by means of which we can represent, categorize, configure and comment upon our experiences. Such resources allow us to distance ourselves from the world, and at the same time we become able to perspectivize what we see and hear *as* something, as instances of something we are already familiar with, or as novelties with some interesting, previously un-noticed, qualities. In other words, representations serve as resources for communicating and meaning-making, and they are essential to all human practices including perception, remembering and thinking and other psychological activities. What we refer to as cognition are activities of meaning-making made possible largely through the use of representations.

In the sociocultural tradition (Wertsch, 1985, 2007; Vygotsky, 1986), the idea of mediation and mediated action is fundamental. The concept of mediation was originally introduced by Vygotsky to argue against the simplistic notion of human behaviour as simple responses to stimuli in the outside world. The idea of stimulus–response connections as the core of human learning and development was the basic assumption of the Pavlovian reflexology (and, in the West, of Watsonian behaviourism) which dominated psychology from the 1920s and during the decades to follow. To Vygotsky this kind of reductionism was a significant element of what led him to talk about the crisis in psychology and the failure of psychological research to contribute to the understanding of the interrelationships between culture, individual activities and what he referred to as "higher psychological processes" (Vygotsky, 1978) (i.e., learning, remembering, aesthetic experience and voluntary forms of behaviour). If we want to understand such genuinely human phenomena, psychological investigations must focus on the interaction between the child and significant others in the sociocultural environment; mental processes are social in origin and they are mediated through interaction.

Mediation implies that humans, unlike other living creatures, think and act in "a roundabout way" (Vygotsky, 1994, p. 61), that is, by means of cultural artefacts that have evolved over

time – symbolic representations (including linguistic tools) and artefacts. We think and act in social practices by means of concepts such as triangle, centimetre and democracy. In a similar vein, we use physical artefacts such as kitchen utensils, telephones and a range of other prosthetic devices in our daily activities. This distinction between artefacts and intellectual tools (the latter Vygotsky referred to as psychological tools, cf. Kozulin, 2003), by the way, is not to be taken as a categorical difference between a material and a non-material, ideational world (Cole, 1996). On the contrary, we use the alphabet (an intellectual or symbolic tool) and paper and pencil (artefacts) to write, and nowadays many of us use a global postioning system (GPS) navigator loaded with maps and symbols to take us from A to B in an unfamiliar environment. *Re*-presentation is achieved by combining symbolic tools and physical resources. On closer inspection, mediation always, at some stage, relies on materiality.

The tools for mediating the world that we encounter have their origin in society and in human activities; they emerge through practices that extend over time. All readers of this text know how to use expressions such as per cent, circle and virtuality in an intelligible manner in conversations. Through socialization, we appropriate the symbolic tools that are functional for meaning-making in social interaction. We may not understand them fully or know their origin and development (i.e. their sociogenesis), but we know how to use them "for all practical purposes" as ethnomethodologists put it. It is by appropriating such cultural resources that human beings are socialized into communities inhabiting, and continuously developing, symbolic universes, modes of expression, inscriptions and material artefacts. "As children are drawn into cultures, "what is to hand", becomes more and more that which the culture values and therefore makes readily available" as Kress (1997a, p. 13) puts it.

Multimodality, language and activity: a sociocultural interpretation

In the context of the emphasis on considering issues of multimodality in human communication expressed by many scholars (Jewitt, 2005; Kress, 2003b; Kress and Van Leeuwen 1996/2006), there is a reaction against the dominant role that language and discourse play in studies of meaning-making and, perhaps, in studies of cultural practices in general. This critique is fuelled by the current developments in digital technology, where other modes of mediating events and communicating emerge as prominent in many contexts: "we make signs from lots of different 'stuff'" and not just through language, as Kress (1997a, p. 13) puts it. In the sociocultural tradition there is a strong emphasis on the decisive role that language (spoken and written) plays as a symbolic resource. Thus, there is an apparent contradiction between the "multimodal programme" and the sociocultural tradition in the attempts to study meaning-making and cultural production and reproduction.

We would, however, like to preface the presentation of some examples from our own work on issues of multimodality by commenting on these differences. From a sociocultural perspective, the significance of other forms of mediation is readily accepted. Some of the earliest forms of symbolic expressions that we know in history are artefacts such as rock paintings, engravings on tools and ornamental jewellery to be carried in ceremonies, i.e. resources for communicating that were not directly based on linguistic mediation (although we would argue that, most likely, meaning-making practices involving conceptual mediation and analysis must have played a role when they were produced; but this is another matter). In fact, it is not until late in history that signs that are connected to linguistic (and numeric) systems appear, or, in other words, that inscriptions began to emulate speech. Put differently, in the sociogenesis of symbol systems and signs, images preceded representations of linguistic units such as words, sentences and phonemes; people represented what they saw or imagined (Lewis-Williams, 2002) long before

they began writing words and well before they began to deconstruct words into smaller elements (Olson, 1994).

It is also obvious that other forms of meaning-making are significant in human development. Drawings, pictures and other symbolic tools are not second-order representations to language, but rather important elements of the human repertoire for meaning-making and for the building up of a social memory relevant for specific practices. One obvious example of this is the development of the techniques of making charts and maps to be used in navigation. The particular manner in which we now mediate the world through such artefacts has a long history which includes the development of projection techniques, the design of conceptual constructions (such as latitude and longitude) and even political deliberations and considerations (such as, for instance, no longer making Jerusalem the centre of the world map) (cf. Williams, 1992).

There is, however, some sense in which verbal language holds a special position in the socio-cultural interpretation of mediation. Indeed, Vygotsky's famous dictum of language as "the tool of tools" testifies directly to this central role ascribed to linguistic mediation. Through the origin of this perspective in developmental psychology, a fundamental question is that of the connection between the biological, social and cultural dimensions of human development. The problem can be understood as a matter of how a biological creature such as a newborn child appropriates the cultural tools of his immediate and, eventually, more distant environment, to become a full-fledged participant in social interaction, equipped with a range of interactional and semiotic skills. In these processes of socialization, or, to put it more bluntly: in this process of humanization, language plays a fundamental role. In the words of the evolutionary psychologist Merlin Donald (2001), culture "hijacks" the human brain largely by infiltrating it with cultural categories and ways of understanding the world, and the most important mechanism by means of which this is achieved is language. In endless communicative encounters, we learn to structure and categorize what we hear and see by means of linguistic resources, and learning and development result in an increasing linguistic and symbolic control of our experiences (Rommetveit, 1985). In mundane, everyday conversations and interactional practices we *re*-present the world and *re*-call our past experiences and make them relevant to our present concerns.

Thus, with respect to the issues of what constitutes a mode, and in what ways modes differ, we would like to suggest a view where speech, writing, images, graphs and other symbolic resources are seen as elements of human repertoires for externalizing and objectifying experiences and for communicating about them (Säljö, 2005). These forms of mediation are interrelated, and they interact in complex ways in almost all kinds of social practices. Thus, our perceptions of a painting are shaped by discursively mediated ideas and conventions. We understand why the works of Picasso or the impressionists are interesting and innovative, since we know something about how they challenged contemporary traditions of painting. In a similar vein, our readings of a map of the world are informed by conventions and canons of representation (Wartofsky, 1979, 1983) for how one makes a projection of the world in two dimensions that have evolved over centuries. In more specialized activities in society, nurses and physicians learn how to read X-ray images and images from PET-scanners by being taught by more experienced colleagues what there is to "see" in what is represented. This learning is embedded in visual and discursive practices, where the two modes build on, and presuppose, each other in a successive shaping of the abilities of the novice to single out what is relevant to attend to (Ivarsson and Säljö, 2005).

Illuminating contemporary studies of such shaping of visual perception have been carried out by Goodwin (1994, 1997, 2001) in the context of doing science. What Goodwin shows in one study is how archaeologists engage in a particular kind of "seeing" when looking at objects and artefacts when excavating. In a different study, he illustrates how indicators in scientific measurements rely on users' ability to identify quite subtle variations in what there is to "see" when

looking at a specially prepared piece of string serving as a test of chemical reactions in water. In the latter case, Goodwin (1997) demonstrates how perception has to be shaped in order for a less experienced member of an exploration team to see something as a particular kind of black, and to distinguish the relevant kind of blackness from other kinds of black. This shaping of vision in which one learns to discriminate between "black" and what is referred to by the experienced members of the team as "jet black" takes place through situated practices where language plays a crucial role as a mediating tool when observing artefacts.

In a totally different setting, these intimate interrelationships between speech, writing and images as resources for communication in human practices have been documented in the fascinating work by the archeologist and art historian Schmandt-Besserat (2007). What Schmandt-Besserat shows in her studies of the sociogenesis of literacy in the ancient Near East is how "writing shaped art" (p. 15) in the sense that pictures and images on pottery, glyptic art and on wall and floor paintings changed in character and composition as literacy spread as a mode of communication. Basically this transformation implies a shift where "art benefited from the paradigm developed by writing to communicate information" (p. 25). Thus, in the "late fourth millennium B. C., pottery paintings left behind preliterate repetitious animal motifs and glyptic abandoned its circular topsy-turvy compositions in order to tell stories" (p. 102). While "preliterate painters tried hard to achieve the utmost stylization, those of the literate period aim to be as informative as possible" (p. 24), and they attempted to tell stories by means of images. A psychologically interesting consequence of this development is that "preliterate pottery compositions formed an over-all pattern meant to be apprehended as a whole, or *globally*", while "those of the literate period were to be viewed *analytically*" (p. 24), i.e. as portraying events, dramatic episodes and social relationships that emerged as the various elements of the composition were viewed in relationship to each other. At the technical level of producing images, this implies that elements such as the design of space, the location of characters and their size and relative position began to be used as indicators of importance, hierarchical relationships (size as an indicator of rank), status and so on in order to narrate events. Thus, in "the literate period, 'reading' images became akin to reading a text" (p. 25). Or, in other words, the linearity of text and narrative began to have implications for the manners in which images were organized.

What is interesting for our purposes in Schmandt-Besserat's in-depth account of the intricate and dynamic relationships between image-making and writing in this formative period of the development of resources of objectification of human experiences is how the two visual modes interact with each other and assume different, and often complementary, functions. Thus, not only did written narratives provide models for art, there were also consequences in the other direction. Inscriptions on art objects, such as names and other messages, could take on a life of their own when presented in the context of a stele or a statue. The inscriptions would do one kind of job (for instance preserving the name of the king or the deity for eternity), while the material artefact would provide other symbolic information regarding the status of the person, his wealth and other qualities (for instance through its size or the nature of the material used for making a statue), Thus, the different symbolic elements of a statue such as its size, the material used, the posture of the person or deity portrayed and the inscriptions all contributed to the meaning-making.

In the following we want to approach the issue of multimodality from a sociocultural perspective by means of a few examples intended to illustrate the interdependencies between modes. The point of our examples is to argue for the value of studying representations when they are part of human practices in which meaning-making is contingent on what, in Kress' terms, "is to hand" through different modes. Thus our point of departure to some extent follows Wittgenstein's (1953) dictum of studying meaning-making not in the representations

themselves (be they words, images or something else), but rather in how they are used in social practices within activities and how they interact (Mäkitalo and Säljö 2009).

By means of three brief empirical cases, we will illustrate how discourse is integral to meaning-making in multimodal settings, i.e. the point is one of showing how language fulfils a bridging function when engaging in multimodal communication. The first two examples are taken from contexts of gaming, and the third one is from a particular kind of practice that is central to the training of architects at university. Our ambition is to show how concepts emerge in practices as situated responses to what is happening in a world of non-linguistic representations. We also want to argue that the emergence of such concepts may be followed by a process of canonization and institutionalization, which eventually may result in a situation where a particular way of speaking and structuring the world will be seen as a more or less stable element of a practice that will be shared by those engaged in it. Once these conceptual innovations have been sedimented and become part of the established practices of meaning-making, new members approaching the specific knowledge community need to master them in order to appear competent. This is the cumulative element of language as a tool for meaning-making.

Illustration 1. Gaming and the invention of concepts

The first illustration of this dynamic interplay between modes comes from the practice of playing computer and video games. Such digital games are commonly described as examples of interactive media with a high degree of multimodality in the sense that they build on combinations of images, animations, texts and sound. However, at the same time such games, when viewed as practices that people engage in, can also be conceived as social environments temporarily inhabited by players who interact with and against each other in order to achieve some built-in goal. When studying such gaming as a practice that people engage in, it becomes evident that discursive activities play a crucial role for the players' ability to play and to co-ordinate their actions. For instance, gamers can be seen to invent concepts and categories that they need in order to talk about the problems at hand in this multimodal practice (Linderoth, 2004). In the first brief excerpt, two children, Elin and Bea, are playing an action game in a shared virtual world. They are sitting next to each other in front of a television set. The screen image is split in two halves so that each player can see the game world through their avatars perspective. When we enter the discussion in Excerpt 1, the two children have had some problems finding each other in the game world.

Excerpt 1

1. *Elin*: Are you still in the sand-path?
2. (*Bea's avatar has just entered a tunnel with a floor made of sand*)
3. *Bea*: Yes
4. (*Elin's avatar approaches the entrance to the tunnel*)
5. *Elin*: Bea stay there, I will come to the sand-path now, stay there.

From this event and throughout the entire game session, the children make use of their self-invented concept "sand-path" in order to find and keep track of each other. This expression becomes a mediating tool for co-ordinating their movements in the game. By using the "sand-path" as a landmark, they can find each other when they need. Later on in the game session, the children make up another concept.

1. *Elin*: These guns are not good, they are useless, right?
2. *Bea*: Yes
3. *Elin*: Aren't they bad, they are so useless.
4. *Bea*: Yes but not that spin-weapon, did you think that one was bad?
5. *Elin*: No, no the spin-weapon is awesome.

Here the girls are commenting on the qualities of different game weapons and come up with the concept of a "spin-weapon" in order to refer to a certain kind of science-fiction weapon appearing in the game. Both these examples illustrate how gaming as a practice is highly conducive to, in fact sometimes forces, the players to create new concepts as mediating tools in order to accomplish a shared understanding that enables them to go on with what they are doing. Purposive action in the animated virtual world where co-operation is assumed to take place relies on co-ordination by means of language.

Illustration 2. Backspawning and hillhugging

The two examples above show concepts or categorizations that are very local and with a low degree of complexity. They are used by the players to localize each other in this specific game and to co-ordinate further action in a fairly immediate sense and with instant feedback. But the development of such new concepts can of course be more complex and general, and they may point towards more generic sets of behaviours or courses of action that apply to a wider variety of games. Many non-gamers have probably experienced that gaming is embedded in a particular kind of "lingo", a social language of its own to use Bakhtinian parlance (Bakhtin, 1986), and they may have heard gamers use terms and expressions that are incomprehensible to outsiders. From a sociogenetic perspective, this type of game discourse may have emerged from the type of game practices we observed Bea and Elin engage in. Through sustained interaction in online gaming environments, there is a setting in which local conceptual innovations may live on to become permanent parts of an established tool-kit of communicative resources. As an illustration, one can look at the discourse practices used in the context of a battle game such as *Enemy Territory Quake Wars*. In this online multiplayer game, terms such as "backspawning" and "hillhugging" can be heard in voice-chats and read in the game-chat. To backspawn implies to let your avatar reincarnate at a point in the game world that is further away from the frontline; something that in certain situations can be beneficial for your team. "Hillhugging" refers to a type of behaviour where a player hides in the outskirts of the game world and tries to snipe his opponents, even though there are other more important tasks to attend to for the team. Hillhugging, therefore, is not appreciated, and in the game you may hear and read expressions like "work as a team, no hillhugging here or you will be kicked out".

That concepts and categories emerge from local practices, and then become sedimented and canonized within a certain discursive activity is, following a sociocultural perspective, not unique for gaming; rather, this is an illustration of how language develops in the context of new activities and in response to other modes of representation. Since gaming is a rather new practice, this process becomes visible. In terms of how to theorize multimodality, these examples illustrate how actions in the highly graphical game world are constantly mediated through discursive practices; when interacting, and in order for there to be a game, the players constantly verbalize what they see in order to co-ordinate their activities and achieve goals. At an abstract level, images, animations and talk may be conceived as different modalities, but when engaged in practices, the distinction becomes blurred; talk is a mediating tool for acting in the graphical world, but it does its work by profiting on what is available through other modes.

Figure 15.1 Architectural critique as a communicative practice.

While our first two illustrations show a practice in which the participants themselves need to invent some of the terminology they use, our third and final illustration shows a more common pattern. Here the participants are expected to acquire skills and terminology that are already established parts of a social language, i.e. they take over ways of narrating, thinking and producing and using images that are constitutive elements of highly specialized institutional practices. The particular context we will use as an illustration is learning some of the skills that go into reasoning and working like an architect.

Illustration 3. Seeing the cityscape and seeing by means of the concept of cityscape

In our third empirical illustration, two parties are intensely involved in discussing an image within the context of education of architects at university (Figure 15.1). The specific situation is the so-called critique[1] (also known as the "crit" or design review), which is an important element of the training of architects. In this institutionalized practice, the students present their finished projects to an audience of professors, peers and experienced practising architects called upon to participate. The projects are generally, as in the present case, reported as an assembly of images, texts, floor plans, sections, elevations and other projections, which are printed on a few A3 sheets and posted on a wall. The posters, then, function as the material ground from which the students will receive feedback and critique on their work from the members of the audience. Thus, this is a formative moment where the students document their learning, and where they have to defend their solutions to the various dilemmas that characterize architectural design.

In the example used here as an illustration, the students had to solve the task of designing a number of ateliers. These should be placed on a specific building site, a real place in the city of the university. Just before our empirical illustration begins, the student, Anna, has finished the presentation

Figure 15.2 The student image discussed.

of her project. One of the teachers, Bill, begins his critique/commentary. He does so by focusing on one of her images. The image in question is a watercolour painting made by Anna (Figure 15.2), but what is important to remember here is that we, as researchers, are not interested in interpreting the image as such. The work of interpretation, elicitation, discussion, evaluation or whatever we want to call it, is a practical concern for the participants we have studied; this is what *they* are expected to engage in. And it is by looking at *this work* (rather than the image as such) that we can begin to understand the role of the image in this specific practice of training of architects.

Through their commentary work, the teachers and the professional architects display a number of normative judgements and practical considerations in relation to both the educational and the occupational practice of architecture. In this commentary, one important move made by Bill is to introduce the terms/concepts of cityscape and landscape, respectively. By looking at the particular manner in which the notion of cityscape is elaborated and positioned, in relation to the image made by the student, we can begin to understand what makes this a case of architectural education.

Bill: Then this image. It's actually rather thinly sown with images of this kind because in my opinion this is an image that, perhaps not primarily talks about what does my house look like. But instead what does my atelier look like in the cityscape, in the landscape. You know how does it look when I stand and hang around here somewhere (*pointing to where the man stands*), right? And gaze towards Queen's Bridge. Then you will see, that is, you hint at how your buildings interplay with the jetty and, further out one sees the sea, Queen's Bridge which frames it all. So this is an on-the-spot account of the landscape, in which the building is a part, and I think it is a great image that tells about one of your most important thoughts, I believe, when it comes to the placing of the buildings. Isn't that correct?
(*Short pause. Bill, smilingly, makes a gesture inviting Anna to respond*)

Anna: *it could be* (*laughs*)

Bill informs his students that his reading of the image is structured by the concept of cityscape, but he also shows, in and through his commentary, how such a reading should be performed.

In his comment, the importance of showing the exact look and feel of the suggested building is downplayed in favour of other aspects. What is subsequently exhibited as relevant noticings concern how the proposed buildings relate to some of the major landmarks in the scene, such as the bridge and the more distant sea. The relation between the buildings and their foundation, in this case a jetty, is also mentioned, and it is talked about as an "interplay", which is hinted at in the image.

In Bill's reading, this image is indicative of interesting thoughts on how to best place the buildings on the assigned building site. The placing of buildings is a most central and practical consideration that architects have to deal with. This is a multidimensional task that among other things includes an important rhetorical element. The issue is not just about how to place the buildings, but, equally importantly, the suggested placement must be argued for in relation to clients, municipal officials and other stakeholders, who will decide on projects on the basis of this type of presentations. What Bill sees in, or rather by way of, this image is something that is not present in the image itself. Although the jetty represented in the image is part of the building site, there was also a large piece of land that could be used by the students for placing the buildings. For Bill then, who has access to this information, the deliberate choice not to place the buildings on land is thus visible in the talked-about rendition. In other words, what is "seen" is something that is not in the image, but which still is a highly relevant element in how the problem has been solved by Anna.

Furthermore, the way the image is composed is simultaneously seen and commented upon as a constitutive part of the argumentation for this choice. As a professional architect, Bill is in a position to recognize qualities of a project without having them described for him. Those qualities are, in and by themselves, seen as motives for why a student has followed a certain design path. In this particular situation, which is part of a learning context, the unique qualities of the proposed ateliers are perceived and explicitly verbalized by the teacher. But the *verbalization* here is a pedagogical move, aimed at the larger audience of students present as he speaks. The 'crit' is a public event, where one of the pedagogical goals is to provide demonstrations of architecturally relevant ways of looking at (student) projects. What is demonstrated in this example are central elements of the skilled vision in relation to cityscapes. Bill's comment is recast as a question at the end; nevertheless, his remarks about the student's thoughts about placement should not primarily be seen as a genuine question about the presentation. Rather, his extended comment is intended as a model of how one should act and reason as a professional architect. In other words, his commentary is occasioned by the fact that he is acting in a specific context where he can demonstrate to all those present important features of how to analyse images and how to provide solutions to architectural problems.

It is clear that the discussion that takes place in this example builds upon the image that has been produced by the student. The image as such holds a large potential when it comes to possible ways it could be made relevant in the interaction; in fact, the meaning potentials (Rommetveit, 1974) of this image are countless and relative to what activity it is part of. Given a context of arts education, it could for instance be discussed in terms of *technique* (perspective), *choice of material* (pigment behaviour), *aesthetics* (colour balance, temperature) and so forth. But, and this is important for our argumentation in this context, these notions are *not* brought up by the teachers; not in the short excerpt that we have shown, and not in the subsequent interaction.

It is in the selection and display of what counts as relevant ways of talking and thinking about the representations that the training of a specific profession is manifested. In our case, this socialization is in part built upon the practice where professional architects respond to representations already produced by students and when they tell them what those objects *are*, not in any essential sense, but in the *eyes* of the profession. Whether or not the notions of "cityscape", "placing"

or "interplay" were consciously present with the student as the image was produced is made less important. It is the pedagogical practice of meaning-making in relation to the students' products in the particular instructional format that is of prime concern. What is exploited in this particular piece of interaction is a potential of showing how architects reason when they design, and the audience is not just the student who is responsible, but, rather, all those present in the room. The image becomes a vehicle for making architectural reasoning publicly visible. Thus, a particular social language is invoked and sustained as a relevant way of analysing images in the specific practice that is intended to convey insights into how one thinks and makes images in the context of architectural design.

Conclusions: Images, language and social practices

The point of our discussion has been to argue for the interconnectedness between representational tools. Images, pictures and other visual tools form part of the human repertoire for sense-making, and they are embedded in discursive practices both when produced and when read. In order to make a drawing of an object, one has to have a certain conceptual control over how a three-dimensional object can be reproduced in two dimensions. In the evolution of human symbolic representations, achieving this kind of control was one of the dramatic steps towards using images as tools for representing reality, as has been argued by Lewis-Williams (2002) in his analysis of the differences between, for instance, body-painting and the production of symbolic images: the "notion that an image is a scale model of something else (say, a horse) requires a different set of mental events and conventions" (p. 181) than those that went into earlier, pre-historic forms of aesthetic expression, for instance in body-painting. In a similar vein, interacting with new images often requires conceptual adaptation and innovation as we tried to explain on a small scale with the excerpts from the young girls playing computer and video games. Engagement in pictures and animations may very well provoke linguistic innovation and *vice versa*.

A second point that we have argued for is that meaning-making is always relative to social practices. What is a relevant way of perceiving an image cannot be decided on, unless one considers the practice of which it is a part. All representations have meaning potentials, but what aspect of those potentials that are exploited is a situated affair? As we have tried to show with our brief example from the specific context of the training of architects, the images produced are subjected to highly specialized forms of analysis that are intended to communicate to students features of what Goodwin (1994) refers to as professional vision; institutionalized ways of "seeing", reasoning and producing images.

As Goodwin argues in his seminal paper, we see through the conceptual world of language, and this also goes for the perception of images. Even though images have some unique properties for meaning-making, for instance in terms of spatially organizing information, they are not appropriated in the same sense as are concepts and terms. Once a concept is appropriated, it becomes a tool for seeing and for extending our agency in a certain practice. From this point of view it will be less productive to separate images and objects, or even man-made items and natural objects. Be it a tree, a house or a painting, once our mind is "hijacked" by a certain discursive practice or social language, we will perceive the world through cultural lenses. But a fruitful angle from which to study the development of our capacities for communicating through different modalities is to take one's point of departure in practices rather than in qualities inherent to words, narratives, images or other representations. Wittgenstein (1953) advised philosophers not to analyse language when "it goes on holiday", as is so characteristic of traditional analytical philosophy when asking essentialist questions of "what is X?". Rather, we should

investigate how language is used in interaction by people in situated activities. We believe that to understand human development and learning in the context of the developments brought about by digital media, a similar attitude has its merits; it is in social practices where people make use of increasingly diversified cultural tools that we can begin to explore what cognitive and aesthetic socialization is all about in contemporary society. Words, narratives, images and other representations may be distinct as cultural entities, but in meaning-making in practice they are always contingent on each other.

So, what does a sociocultural approach have to offer research on multimodality and multimodal practices, and where are the shortcomings? The basic unit of analysis in sociocultural research is the notion of mediated action. This implies that the link between people and cultural tools is always part of the inquiry; people always act by using making use external resources. Such a perspective, thus, is an alternative to more essentialist approaches which analyse language, images or any other form of representation as disconnected from other resources for meaning-making. The other element of this equation is the emphasis on the situated nature of human action and communication; activities are always embedded in activity systems. The same image, for instance a drawing of house as in the case of the training of architects, can be part of many different activities, and people orient to it accordingly. If the analyses are carried out properly, situated understandings of the uses of cultural resources will emerge. An important component of such an approach is that it attempts to retain the integrity of human activities as they unfold in an increasingly complex and diversified society.

It is in our view important to conceive of the sociocultural tradition as a perspective or an approach to research on communication and multimodality. Thus, it is not a single theory in any strong sense. Rather it invites consideration of how people engage in mediated activities and how they learn to interact and represent the world by means of the resources that they are accustomed to. The insights provided thus offer a complementary picture of what human development and learning are all about, and in this account the mastery of different modes for representing the world is one the most exciting features of individual and collective development in contemporary society.

Further reading

Daniels, H., Cole, M. and Wertsch, J. eds. (2007) *The Cambridge Companion to Vygotsky*, New York: Cambridge University Press.

Gee, J. P. (2003) *What Video Games Have to Teach Us About Learning and Literacy*, New York: PalgraveMacMillan.

Goodwin, C. (1997) 'The blackness of black. Color categories as situated practice.' In L. B. Resnick, R. Säljö, C. Pontecorvo and B. Burge, eds. *Discourse, Tools, and Reasoning. Essays on Situated Cognition*, New York: Springer: pp. 111–40.

Stevens, R. and Hall, R. (1998) 'Disciplined perception: Learning to see in technoscience.' In M. Lampert and M. L. Blunk, eds. *Talking Mathematics in School. Studies of Teaching and Learning*, Cambridge: Cambridge University Press, pp. 107–49.

Wells, G. (1999) *Dialogic Inquiry. Towards a Sociocultural Practice and Theory of Education*, Cambridge: Cambridge University Press.

Acknowledgement

The research reported here has been funded by the Swedish Research Council and the Knowledge (KK) Foundation. The authors are members of The Linnaeus Centre for Research

on Learning, Interaction and Mediated Communication in Contemporary Society (LinCS). The chapter was written while the third author was a Finland Distinguished Professor at the Centre for Learning Research, University of Turku.

Note

1 In Swedish the term used in the training is "kritik", which is equivalent to critique in English. However, the intention of this kind of instructional format is not to criticize in any negative sense, but rather to provide expert commentary on the work done by the students.