Some Suggestions for Expanding the Conceptual Framework for HCI Development and Evaluation

Nils Dahlbäck

SICS
(Swedish Institute of Computer Science)
Box 1263
S-164 29 Kista, Sweden
nilsd@sics.se

Department of Computer and Information Science
Linköping University
S-581 83 Linköping, Sweden
nilda@ida.liu.se

ABSTRACT

With the development of computer systems into more and more areas of human life, it becomes possible for the HCI research community to use a wider range and make wider use of concepts from the humanities and the social sciences than has hitherto been the case. In this paper three such conceptual dimensions are presented and discussed. On the level of the individual: cognition-emotion-volition; on the size of the social unit: individual-dyad-group, and finally levels of communicative intentions. The potential advantage of using these conceptual dimensions is illustrated with examples from the fields of personal and social navigation, and multi-modal dialogue systems.

INTRODUCTION

The world of computers and of human-computer interaction is changing rapidly. From the number cruncher to the stand-alone office tool to the networked communication system. From a tool for work to an engine for games and fun and a means of keeping communities together, and even today a world where new communities are created. As long as the computer was a one-person tool for 'serious' work, the theoretical framework for evaluating the tools, and for describing the psychological aspects of the work, could be rather simple and one-dimensional, focusing primarily on 'efficiency'. But with today's strong wind of change, and with clear signals that it, if anything, will grow stronger in the times to come, such a one-eyed focus is not enough (and it probably wasn't really enough before either). The aim of the present paper is to suggest the outlines of a possible framework for analyzing and evaluating today's systems. I will not attempt to describe in detail the situation today, but rather give a possible bird's eye view of where we are and where we could be heading. I am well aware that with such a distance to the object viewed, there is of course a risk that what is seen is a rather distorted picture of what is really there.

The basic argument is very simple. When computers are entering into more and more facets of our life, we should incorporate into our work more aspects and categories of human psychology than what have hitherto been the case. This will help us see clearer both what we have today, and what we don't have but might want to develop. The approach I will be using is to mention some of the distinctions I think fruitfully could be introduced into our work, and illustrate the advantages with concrete and specific examples. The examples will be taken from research that I am familiar with (personal and social information navigation, and multi-modal dialogue systems), and hence be unashamedly biased. But my hope is that there will be some generalizability from these examples to also other areas of the HCI community.

NOT ONLY 'HOMO COGITANS'

We are all something more than thinking minds with bodies. We laugh and fight and want and … Already at the turn of the century individual psychology was divided into three major sub-areas; cognition, emotion, and volition (or thinking, feeling and wanting/needing). Which all can be studied separately, as has been successfully illustrated by psychological research for a century. But this is not the same as saying that we at specific times are either thinking or feeling or wanting. If we look at much of mainstream HCI research, however, isn't this the picture that emerges? We are mainly studying tools, and we do that from the point of view of efficiency. (How fast can one perform a task, or learn to use a system, or…) But many if not all of us do not select tools only by how efficient they will help us perform a task. When driving a car from San Francisco to Santa Barbara, most of us would prefer doing it in a BMW than in an old Chevy, even if the driving time would be exactly the same. And some of us are willing to pay 10 times as much for the BMW than for the used Chevy, even if it isn't going to make us perform the task of driving any better, and definitely not ten times faster. Of course part of this could be attributed to status and vanity and similar factors, which, some might argue, should not be a concern for the science of HCI. But this is not all that there is to it. Some tools are preferred simply because they are more fun, or aesthetically satisfying, to work with. And this in turn makes us more likely to want to use it.

The argument here is therefore not that we should also look at the games and other 'fun' software, because people do

not use computers only for work. The argument is instead, that we should *always* have all three aspects in mind; as designers when we are developing systems, and equally so when evaluating them. The Agneta & Frida system (Persson 1999) is an example of an attempt to incorporate the fun in the serious. We can argue about whether it works (and we certainly have argued about it!), but regardless of our position towards this particular example, I think it serves as a good example of trying to break out of the 'efficiency-barrier', and acknowledging that fun and efficiency need not be in opposition but rather something that we should strive for to be there in concert.

NO USER IS AN ISLAND

The think-feel-act dimension discussed above is primarily intra-individual. But of course we are never really isolated individuals, but always members of smaller and larger groups and cultures. There are many different levels here, and all distinctions are not always useful. But at least the following ones are probably necessary to keep distinguished: individual, dyad, small group, organizational, and societal. And this is recognized by the HCI research community, where the social aspects of HCI are ubiquitous these days. Not only in areas such as CSCW and studies of MUD and MOO systems, where the social dimensions are obvious, but also in other not so obvious areas such as information retrieval.

An emerging paradigm for information navigation is 'social navigation', e.g. (Dieberger 1997) (Munro et al. 1999 (forthcoming, August)) which builds on the basic insight that most real life information retrieval tasks are pursued in whole or part through personal and social interaction with information providers. Searching for information in hypermedia is often today seen as similar to navigating in geographic space. The insight behind the social navigation paradigm is that navigation in geographic space is often a social task. We ask for the way, follow people that know the way, etc. A commonly used distinction in this context is between 'direct' and 'indirect' social navigation. In the navigation in the real geographic world, the meanings of these terms are straightforward. Direct social navigation is asking someone for the way to the destination, indirect is by following road-signs or paths. And in navigation in electronic worlds, similar situations can be found. But there are here also cases that are difficult to classify using this distinction. What, for instance, is a recommendation system? It is not the same as being in a dialogue with another person, but nor is it the same as following a path or a road sign. When cases like this emerge, we can either try to squeeze them into our conceptual scheme, or we can start reconsidering the scheme as such. I will here suggest an alternative framework, which, as we will see, will be applicable and have possible consequences for not only this sub-field of HCI, but for the field at large.

ACTION IS NOT INTERACTION

The first distinction I want to introduce is between *action* and *interaction*. As human beings we are living in two worlds, the physical and the social. And our ways of being in this world is based on two different modes, as it were. In the physical world we *act* on objects; we move them, we transform them, we even destroy and digest them. In the social world we *interact* with other people. On a basic biological and even ontological level, these are very different kinds of activities. Perhaps the most important difference is that in the social world, the others are also acting agents, which not only can move on their own, but which also have their experience of the world, including their experience of me. It is only with these other agents that we can communicate, in the strict sense of the word. Communication is, after all, derived from the Latin 'communis' and means hence more or less 'making (something) common'. When moving a stone, I am not communicating.

If this argument is accepted, not only becomes 'social navigation' something of an oxymoron, since navigation is primarily concerned with moving in the physical world. But we are even more or less cornered into a position which entails that most of so-called human-computer interaction is not really interaction, but simply action. In the world of direct manipulation, there is only action. It is my belief that re-conceptualizing what is happening when a person is using a computer is in fact fruitful, and will lead us to look at familiar cases view new eyes. So this is not just a quibble with words.

RE-CONCEPTUALIZING MULTI-MEDIA SYSTEMS

With the distinction above in mind, we can look at multi-modal systems and interfaces as not being one class of systems, but actually two. In one case, we are acting in an electronic space through different modalities, in the other we are interacting with an agent that can use more than one input or output channel. The crucial difference is that in the latter, but not the former, case, we need a dialogue manager (Jönsson 1992), (Jönsson 1997) at the center of the system, which co-ordinates what is going on through the different channels.

Of course the distinction is not as clear cut and simple as I have sketched it above. The categories here are graded or radial (as Lakoff (Lakoff 1987) and others would claim is true for all natural categories). The prototype here is the interaction between two grown-up people sharing the same cultural background. Children are different, as are animals. But isn't it the case that the further we move from these prototypical instances of dead matter and grown up people into the border zone, the more uncertain do we become in whether we should treat what we encounter as belonging to either of these kinds? Isn't the case that this is a zone where intuitions differ also within societies that are rather culturally homogenous? Within computer science the long debate about whether computers can 'think' is a well-

known example. The differing views on e.g. animal rights is another from a field outside computer science.

My own guess, for what it is worth, is that these two different kinds of being in the world, the social and the physical, are at least in our culture rather clear-cut different types, which bring out different sets of expectations in us.

Coming back to HCI after this detour, I would like to put forth the following suggestion. We are, by culture or evolution, not comfortable with meeting something that we with confidence cannot clearly classify into either of these two basic categories. And computer systems can (with some qualifications) be both. Therefore it becomes important when designing systems, to place them as squarely as possible in either of these two categories. If we succeed with this, users will be able to use their intuitions from their non-computer lives when understanding what to expect. If we end up in an in-between position, we risk triggering unrealistic expectations about what can and cannot be expected from the system. At least if we do not make very clear the limits of e.g. the dialogue powers (in the strict sense used above) of the system. That, at least, is the conclusion we have drawn in our work with developing multi-modal natural dialogue interfaces (Dahlbäck and Jönsson 1997), (Dahlbäck 1997). If we cannot build a system that makes use of both pictures and language the way a human can point to e.g. a map as an answer to a verbal question, instead of giving a verbal answer, and which knows that these actions are actually not only actions but interactive moves in a dialogue, then we better make very clear the lack of connection between the different in and output modalities, or we risk confusing the users far too much. And especially for systems for occasional use by non-experts this becomes crucial.

We can perhaps make use of the action-interaction distinction when trying to make sense of the conflicting results emerging from evaluation of anthropomorphic interfaces. Are the users actually interacting with the agents or viewing them acting, and are the expectations of the users different in the two cases.

LEVELS OF COMMUNICATIVE INTENTIONS

The final dimension I would like to discuss here concerns levels of communication, or perhaps rather levels of intentions in communication. Condensing from the work of e.g. Grice in philosophy (Grice 1975), Rommetveit (Rommetveit 1987), (Rommetveit and Blakar 1979), (Rommetveit 1990) and Clark (Clark 1992), (Clark 1996) in psychology, and Allwood (Allwood 1976), in linguistics, we can talk about at least three different levels of communicative actions. The first is simply acting without the intention of communicating anything. It is therefore, in a sense, outside the communicative continuum, but it can be interpreted by an outside observer as 'communicating' something To use an example introduced by Rommetveit, if I move the lawn on a Sunday morning, to me this probably is no communicative act and has no such intentions. The lawn simply needs moving, and I happen to be awake and feel like doing it. But to my neighbor, to whom I am new, since I just moved into the house, this might indicate (or communicate, in a very loose sense of the word) that I am not the kind of person that party all Saturday night.

The next level is intending to communicate something, without intending that intention to be communicated. New, as I am, to this neighborhood, I might move the lawn every Sunday with the intention to be seen as a reliable and trustworthy neighbor that takes care of his property, and which will not upset this nice community in any way. I will therefore be accepted in the community, be invited to the barbecues etc. But note that my intention to communicate this will only work if that intention is not noticed. If it *is* noticed, I will probably be classified as a rather ingratiating person, and my intended goals will not be reached.

The final level is, then, communicating something, and intending that intention to be communicated. It has been argued by workers in the field that these 'nested' layers of intentions are needed to analyze various forms of indirect communication, e.g. indirect speech-acts and irony. One important aspect of this, is that this introduces not only the possibility but the need to be able to reason about the interlocutor's intentions and belief, but also about his or her belief about my belief about her belief etc. (The ways out from the threatening infinite regress that this analysis suggests have been analyzed by e.g. Clark e.g.(Clark and Marshall 1981).)

I will, however, not pursue that line here now, since that would take us too far off our track. Instead I will use the tools described here in re-analyzing the concept of 'social navigation'.

RE-CONCEPTUALIZING SOCIAL NAVIGATION

The first possible observation is that navigation belongs primarily to the physical world. Navigation in geographic space is concerned with finding the way from one place to another, monitoring the path during the process, and recognizing the goal when it is reached. In many cases, what is called navigation in electronic worlds, is comprised of more than this, but also with evaluating and perhaps even using the information reached. (Kim and Hirtle 1995) make a distinction between three aspects of information navigation; navigation work, information work, and the co-ordinating of the two tasks. Viewed in that light, there is not much social in navigation, except for asking for the way or reading the road signs. (I'm excluding here cases where the navigation in itself is a social process, as described by e.g. (Hutchins 1994).) And the 'social' here is of course not unique to navigation. We are always social beings, and all our tasks are done in a social context. And since we do not talk about 'social car repair' or 'social cooking', why should we talk about 'social navigation'?

But there are many aspects often discussed under the heading social navigation that do not fit into this restricted

view of navigation, e.g. evaluating the information found by judging its source (is it a reliable person or not) or by having the information adapted to the questioner's particular needs as clarified in a dialogue. There is no denying that these are social processes. But perhaps they could fruitfully be viewed as just social processes, without the 'navigation' prefix? Using the concepts discussed in this paper, we could then re-analyze the 'indirect' Vs 'direct' navigation as follows.

Examples such as paths through a forest are not communicative acts. They are similar to the moving of the lawn in the first example above. There is no intention to communicate here, although most of us will interpret them as signs of people having walked there before us. But this is closer to seeing the dark clouds above and predicting a coming thunderstorm than any communicative move.

The reading of the road sign is communication. It is not indirect, though. The distinction between the road sign and the answer to the question from an informant is that the road sign is a kind of mass communication to a generalized receiver. And the important difference between this and the answer to the question from the man in the information booth is that the latter is adapted to the particular individual asking the question. The difference is therefore, in a sense, not about directness or indirectness, but rather between communication or social interaction on the level of the dyad or the group or community.

My suggestion here is that there are two advantages with this analysis, compared to the direct-indirect social navigation conceptualization. First, it opens up the possibility of using results from decades of research on human communication in analyzing and conceptualizing the different kinds of information navigation. It is the 'don't invent the wheel again'-argument once more. Second, we can describe and discuss the difficult cases with more ease. Take the recommendation system, for instance. It can be conceptualized as a social interaction on the dyadic level between a person and a computer agent, which tries to answer, or rather give suggestions, in a way adapted to the particular user's needs. And viewed like this, we can once again use what is known from social science research on factors affecting the uptake of information. How important is it, for instance, that the information provider can give reasons for the particular suggestions?

This example illustrates also how computer systems performing social tasks introduce new questions. Today's agent based systems can on the one hand be viewed as synthetic or artificial agents performing tasks ordinarily done by human agents. But quite apart from the lack of proficiency compared with a human agent for most of those tasks, there is an obvious difference in the level of communicative competence. We might want to claim that the systems have intentions to communicate. And it could have a user model. But no systems that I am familiar with (except for some research prototypes) can have a model of the user's model of the system's beliefs etc. How will this affect the communication? Probably not much, until there is a breakdown. It is when trying to resolve those that the nesting of levels comes into play. The important point is not that we can predict that the agent based system will be worse off than a human when there is a communication breakdown (we all knew that beforehand, anyway), but rather that we see the need to find an answer to the question how this will affect the users' perception of the reliability and usefulness of the system's help.

MUCH ADO ABOUT NOTHING?

But didn't we all know this before? Have I really said something new? Well, yes and no. We can all point to examples of research where the aspects illustrated here have been addressed. So in this sense there is nothing new here (with the possible exception of the distinction introduced here between action and interaction). But doing it, and seeing what one is doing within a larger framework is perhaps not exactly the same thing. My personal belief is that a wider and at the same time more precise theoretical framework will be of benefit to the HCI community, both by guiding the design and by guiding the evaluation of the systems we build.

But here, as always, we need to steer between Scylla and Charybdis. The example of social navigation discussed here also illustrates the other side of the coin; that creative work can start from a position that an outside observer might view as conceptually flawed. Consequently, it is of utmost importance not to base the evaluation of the value of particular pieces of design, or for that matter, fields of design, primarily on the given reasons for its development. The evaluation of a design can rarely if ever be viewed as an evaluation of the theories that inspired the design, since there is much more of a 'creative gap' (filled by the designers) between the theory and the designed artefact than between theoretical concepts and an experimental setting in empirical sciences.

To continue using the case of social navigation to make the arguments here more concrete and specific, the work within this field has opened up important social aspects of tasks such as information retrieval, and illustrated how these features can be incorporated in the design of systems supporting this. But when taking these intuitions and successful instances of design further, it is my belief that the better the theoretical framework used for analyzing and talking about the domain under study, the better the chances are for the further development of the field. My personal opinion is that the 'social' in social navigation has a much wider application within HCI than just navigation, however wide that concept is defined.

SUMMING UP

The present paper makes a two-layered argument. First, that with the change of computers from basically a work tool to something ubiquitous in most areas of our life, the HCI research community could fruitfully make use of a wider

range of concepts from behavioral and social sciences than what has hitherto been the case. Second, a few such conceptual dimensions were used to illustrate possible advantages of using such a wider theoretical framework. No claim is made that these are the most important ones for such an enterprise. The ones discussed here are the ones I have found useful in my own work, and I hope their usefulness are illustrated in my examples. But I hope the cases discussed at least will

ACKNOWLEDGMENTS

Discussions with the members of the Persona project (SICS, Sweden, and Napier University, Scotland), and with members of NLPLAB at the Department of Computer and Information Science at Linköping University have influenced the ideas expressed here. The inspiration is gratefully acknowledged, but none of the members of these groups should be held responsible for the views expressed here.

Funding for this research was provided by EU i3 long-term research through the Persona project, and by the Swedish Council for Research in the Humanities and Social Sciences (HSFR).

REFERENCES

Allwood, J. (1976). ? Linguistic Communication as Action and Cooperation,? Ph. D., Göteborgs universitet, Göteborg.

Clark, H. H. (1992). *Arenas of Language Use*, University of Chicago Press, Chicago.

Clark, H. H. (1996). *Using Language*, Cambridge University Press, Cambridge.

Clark, H. H., and Marshall, C. R. (1981). ? Definite reference and mutual knowledge.? Elements of Discourse Understanding, A. Joshi, B. Webber, and I. Sag, eds., Cambridge University Press, Cambridge, MA.

Dahlbäck, N. (1997). ? Towards a Dialogue Taxonomy.? Dialogue Processing in Spoken Language Systems, M. M. Elisabeth Maier, Susann LuperFoy, ed., Springer Verlag.

Dahlbäck, N., and Jönsson, A. ? Integrating Domain Specific Focusing in Dialogue Models.? *EuroSpeech-97*, Rhodes, Greece.

Dieberger, A. (1997). ? Supporting Social Navigation on the World Wide Web.? *International Journal of Human-Computer Studies*, 46, 805-825.

Grice, P. (1975). ? Logic and conversation.? Syntax and Semantics 3: Speech Acts, P. Cole and J. Morgan, eds., Academic Press, New York.

Hutchins, E. (1994). *Cognition in the wild*, The MIT Press, Cambridge, MA.

Jönsson, A. (1992). ? Dialogue Management for Natural Language Interfaces -- An Empirical Approach,? Ph. D., Linköping University, Linköping.

Jönsson, A. (1997). ? A model for habitable and efficient dialogue management for natural language interaction.? *Natural Language Engineering*, 3(2/3), 103-122.

Kim, H., and Hirtle, S. (1995). ? Spatial metaphors and disorientation in hypertext browsing.? *Behaviour & Information Technology*, 14(4), 239-250.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things*, Chicago University Press, Chicago.

Persson, P. ? AGNETA & FRIDA: Business *and* Pleasure.? *Workshop on Basic Research in HCI at CHI'99*, Pittsburgh, PA.

Rommetveit, R. (1987). ? Meaning, Context, and Control.? *Inquiry*, 30, 77-99.

Rommetveit, R. (1990). ? On axiomatic features of a dialogical approach to language and mind.? The Dynamics of Dialogue, I. J. Markova and K. Foppa, eds., Harvester Wheatsheaf, Hemel Hempstead.

Rommetveit, R., and Blakar, R. M. (1979). ? Studies of Language, Thought, and Verbal Communication.? , Academic Press, London.