From: HCI-02 Proceedings. Copyright © 2002, AAAI (www.aaai.org). All rights reserved.

Contextual Inquiry in HCI: Lessons from Aeronautics

Sidney W. A. Dekker[1] and James M. Nyce[2]

[1] Department of Mechanical Engineering, IKP, Linköping Institute of Technology
SE - 581 83 Linköping, Sweden, sidde@ikp.liu.se
[2] School of Library and Information Management, Emporia State University
Emporia, KS 66801, USA, jnyce@rocketmail.com

Abstract

Human-centered HCI systems can result if developers pay heed to the orientations, expectations, and understanding of the (end-)users. Contextual inquiry has emerged as one way since it can reveal what (computerized) work means to the practitioners who do it, but it needs to make the jump from the description and analysis of current working practice to a design language targeted at the future. In this paper we use three examples from studies into the use of flight strips in air traffic control for their ability to make this jump, extracting the lessons we still need to learn if we want to employ contextual inquiry as a tool in creating HCI systems.

Contextual Inquiry in HCI

Human-centered HCI systems may result when at least equal weighting is given to the orientations, expectations, and understandings of the (end-)users as to those of the software engineers or developers. This is not unproblematic. Quite apart from the normal design pressures that seem to perpetually turn user-centered intentions into technology-centered systems (Anderson, 1994; Forsythe, 1999), the question of what the orientations, expectations and understandings of the (end-)user actually are is non-trivial. Involving (end-) users in the design process is not the panacea (Giddens, 1991; Singer & Dekker, 2000), nor is an overdose of verification and validation (Woods & Dekker, 2001), nor is "getting human factors in early" in the design process. "Newell's catch" describes the basic paradox of designing human-machine systems: The designer makes predictions about the impact of new technology on an operating world. Testing this prediction ultimately requires fielding the system, but this means that the system must be realized at a variety of levels (specific interfaces, software, training for users, etc.). By this time so much commitment and cost (psychological, organizational, political, financial) has become involved, that the potential for changing the design given the information and feedback is minimized. Dissatisfied with the inability to resolve Newell's catch, more HCI work has become "front-loaded". It focuses on understanding the nature of practice before beginning to develop, let alone fielding, new technology (Norros & Klemola, 1999; Hollan *et al.*, 2000).

With the promise of privileged access to an understanding of what activities mean to the people who do them, designers have sought enlightenment in contextual enquiry (Beyer & Holtzblatt, 1998) and ethnography as a method to do it (Nyce & Lowgren, 1995; Harper, 2000). Contextual inquiry is finding out about people's work *where* they are doing that work, *while* they are doing it, and finding out what doing that work means to *them*. HCI design, by extension, is not so much about the building of artefacts or systems, but about designing new ways to work. The role of context in shaping people's practices, and people's perception of their practices, is deemed crucial. This is in line with the growing ecological commitment of many in human factors engineering: to them human performance problems are not about hypothesized internal mental structures that impose constraints on the processing of information from the world, but rather about seeing how features of the engineered world impose constraints and affordances on goal-directed behavior (Vicente, 1999).

Ethnography Versus Design

Many, especially in CSCW (Computer Supported Cooperative Work), see contextual inquiry as the ultimate road to human-centered HCI systems (Harper, 2000). After all, if you tailor the system to how users themselves see their work and their world, you cannot get it wrong. It has to be the summum of human-centeredness. But designers are not ethnographers—the kind of professionals equipped *par excellence* to conduct contextual inquiry. In fact, most designers don't even *want* to be ethnographers and cannot be bothered to learn all that it takes, especially the analytic part that makes for 'strong' ethnography (Anderson, 1994; Forsythe, 1999). At the same time ethnographers have tried, but almost equally often failed, to provide meaningful input into designers' choices. This is mostly because ethnographers cannot make the jump from the description and analysis of current working practice to a design language targeted at the future (Woods et al.,

Copyright © 2002, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

1996). We could call these "designable futures"—descriptions of (future) cognitive work that allow designers to tailor for that kind of work proactively. Or, for their part, ethnographers cannot be bothered to make the jump to designable futures because it is not in their professional doctrine (Plowman *et al.*, 1995). Harper (2000) in fact denies any role for ethnography in requirements capture.

The deeper issue, however, is this: those involved in design and development (including ethnographers) are frequently tempted to equate that which informants do or tell them with what ethnography is and what *it* can tell them. Confounding informant understanding with ethnographic analysis has implications for the credibility and contribution of ethnography in creating HCI systems. Of course, informants can claim privileged access to their operating world: what they know and can tell us always has to be right in a sense, otherwise they could not carry out their work. Yet this is not the same as ethnography, nor is it the same as guiding HCI (re-) design. Tying ethnography to future design hinges on understanding the features and objects of work—not cast in the unpacked language of an informant, but by extracting, revising and verifying the categories into which informants parse their world. It also hinges on finding ways to "build out" this revised, nuanced understanding into a designable future. Strong ethnography is more than record and replay. Analysis requires the ethnographer to move back and forth between informant understanding (native categories) and analytic senses of the work that designers can begin to read as human-centered requirements for the (future) system.

Three Contrasting Studies

Below we contrast three different studies aimed at creating HCI systems in Air Traffic Control (ATC). The studies were all concerned with the so-called "flight progress strip", its role in controlling practice, and whether new ATC systems could do without it (or, rather, could handle some computerized version). All were contextual inquiries to a greater or lesser extent in the sense that they sought to document the role of the flight progress strip where and while controllers were working with them (or without them) and what role the strips had for the controllers using the artefacts in actual practice. Examining this tiny paper artefact, the three studies form a small part of a body of work that is positively huge (see Mackay, 2000) and spectacularly inconclusive. There is no consensus on whether controllers can safely or effectively control traffic without strips (an acute question as various air traffic control systems are due to be replaced or updated with stripless versions). The three studies are representative of this body of work, and instructive with respect to HCI systems not because they converge on a verdict about the fate of the paper strip as opposed to computerized variants (for they do quite the contrary). They present lessons on how to unconfound informant understanding from ethnographic analysis; on how to use contextual inquiry in the present for the design of the future.

Study I

So what are strips for, really? There is no one answer, and any answer would be so locked into a description of how the controllers themselves see their strips, give them meaning and employ them across operational situations, that it would take an ethnography by itself to get there. This did not deter (Albright *et al.*, 1996). Pulling out an impressive arsenal of human factors methods including direct observation, communication monitoring, biographical questionnaires, time and motion measurements, system performance evaluation and subjective workload assessments, the researchers set out to study how controllers managed traffic without flight strips. The researchers' entry assumption was that flight strips support "information retrieval", and that controllers will have to "compensate" for their removal by retrieving information about flights from other sources such as the radar screen (p. 1). The researchers, however, never left an analytic trace for others to critique of follow. The research does not reveal how it came to "information retrieval" as central category in the first place. Throughout the study, however, the starting category of "information retrieval" was never challenged, or revised, and only superficially unpacked. In fact, results derived from stripless control studies were cast in terms of controller information retrieval too, for example the number of times a controller now "requested information from the pilot" (Albright *et al.*, 1996, p. 5). Results, cast in such terms, showed that controllers could retrieve information elsewhere without negative consequences for their perceived workload or ability to manage the traffic. Replacing flight strips with more automated systems, in conclusion, would be feasible. Although a contextual inquiry in the narrow methodological sense, the entry category imposed by the researchers throughout, pre-ordained the nature of the things they saw, the data they gathered and analyzed. The whole endeavor was profoundly etic (looking in from the outside) in perspective. By never challenging or revising the "information retrieval" category, and by mistaking their own reality for the controllers', the research produced foregone conclusions and interpretations. The flight strip is merely one artifact that supports information retrieval, and an inefficient one at that. It can be substituted by other artifacts, such as the radar screen, and its lack can be compensated for by engaging in other activities, such as asking other participants in the system.

The message from such work is that designers can substitute computer-based artifacts for paperbased ones, they can replace human work with machine work without any further consequences to the larger human-

machine ensemble or human expertise required to make it all work. This is the substitution myth (Hollnagel, 1999). The idea is that computerization transforms the tools people work with and forces them to adapt their practice ("compensate"). In this sense, designers only need to cater for the controller activity "information retrieval" in new ways. Reality, however, has more often been inverted: computerization fundamentally transforms human work, and people in turn adapt the new artifacts so that they fit the demands of actual practice. Indeed, others who studied flight strips (e.g. Mackay, 2000) have seen controllers not only read (and retrieve) what is on flight strips. They saw controllers write on the strips, mark them up with all kinds of symbols in different colors, pass them along to each other, tap them or point at them, cross data out on them, talk about them while holding them, fondle them, stack them, sort them, separate them, displace them, categorize them, shove them back and forth, look at them on the neighbor's console, and more. Removing the opportunity to do all this could disrupt traditional controlling strategies in ways not anticipated and challenge practitioners to develop profoundly new ways of working the traffic. On the other hand, of course, all could be activities associated with retrieving (in some way) information (of some kind). But from study I, we would never know that. Information retrieval was taken as inherently meaningful category, never unpacked or more closely specified, around which they entire study and future system would be built.

Study II

Hailed as the mother of contextual flight progress strip studies (Harper, 2000), the Lancaster University project spent many manmonths on the ethnographic observation and analysis of air traffic controller practice, with a particular focus on the use of the flight strip (Hughes *et al.*, 1993). The grand conclusion of this "celebrated" project (see Harper, 2000) was that strips " are the means by which controllers see and note what is happening, what they have already done, what needs to be done. They are an essential feature of 'getting the picture', 'organising the traffic', which is the means of achieving the orderliness of traffic. The strips and their organisation are a proxy orderliness of the configuration of the traffic flow" (Hughes *et al.*, 1993, p. 133).

Strips help controllers 'get the picture'. Such motherhood insight should not have taken longer to arrive at than half an afternoon spent observing air traffic controllers (see indeed Harper, 2000). But it took months. Further, if strips are *the* means for a controller to know what is going on, then there is no point in automating or developing anything new. Such ethnography is Luddism, however unintended, in a new cloak—and not unique among ethnographers, (see Mackay, 2000). It should be no wonder that designers often think they can do just as well, or better, themselves (Forsythe, 1999). As if to confirm the point, the developers, having lost patience with the ethnographers, presented a set of guiding questions for the ethnographers to answer, so they (the developers) could finally get on with building their thing. Here they are (Hughes *et al.*, 1993, p. 135):

- "what characteristics of the existing manual system are unimportant and need not be supported in a computerised system?
- what are important manual activities which need not be supported in a computerised system because the activities are a consequence of the fact that no computer support is available?
- what characteristics of the manual system must be replicated without change in a computerised system?
- what activities from the manual system may be supported in a way which is different from that used in the manual system?"

The developers not only completely missed the point. They did so in a way that was impeccable and eloquent to a fault; so obvious in its simplicity as to be irrevocable. All the old biases of optimistic engineered positivism were there for the ethnographers to see. The developers can simply substitute computers for paper—just tell them which parts you would like swapped. Conceptually, the questions were no departure from the misguided substitution myth that has governed function allocation for decades. Confronted with a set of questions of such mechanistic developer waywardness, the ethnographers never quite recovered. They were unable to formulate a meaningful reply in the remaining pages of their paper and remaining months of their project, and the entire ethnographic-cum-design effort fizzled out on the backburner of mutual misunderstanding. It migrated outside the fringes of techno-push development, resulting only this year (2002) in an ATC system that was hugely over budget, immensely over time, and entirely without flight progress strips.

The failure of ethnographers and developers to get along or even understand each other is not at all unique to this study. It did not have its source in the developers' imperative desire to, well, develop the system. Though emic in perspective (looking out from inside the native), the study presents a particularly naive form of ethnography. Such ethnography does not interrogate that which is common sense to the native. It takes practitioner categories as canonical and inherently meaningful. Of course, informant competence, as expressed in their own words, is strong and valid. But confusing it with contextual inquiry or analysis does not lead to strong ethnography. It also does not lead to design guidance.

Indeed, Beyer & Holtzblatt (1998) explicitly prohibit the use of native categories in design discussions, lest these characterizations trap developers into their entry assumptions or lure them into believing that native motherhoods (such as "flight strips help me get the

mental picture") can actually double as informative analytics. Tellingly, the title of the paper on study II was "from ethnographic *record* to system design", not from "ethnographic *analysis*...". Clearly, a mere ethnographic *recording* does not cut the contextual mustard. The jump from ethnographic record to system design is too large, conceptually and analytically, to be left for the designers to deal with. As the study shows, designers cannot and do not want to deal with it—indeed it should not be their job. Talking to designers meaningfully requires the one who does the contextual inquiry to engage strong ethnography, strong particularly with its analysis. Only such higher-order analytical work can lead to designable futures. The question is: how?

Study III

Informant remarks such as "flight strips help me get the mental picture" should serve as the starting point of a contextual inquiry, not as its conclusion. But how can we move from native category to analytic sense? The revision of categories is a hallmark of strong ethnography, and Ross (1995) (Study III here) has the seedlings of a good example. Surprisingly, it did not do the kind of all-encompassing data gathering of the previous two studies (a massive array of human factors studies or many manmonths of ethnographic observation). It was a simple survey, derided by many ethnographers as an instrument that imposes the researcher's meaning on data rather than bringing out the native's (e.g. Hughes *et al.*, 1993). But Ross (1995) is characterized by the strongest analysis and synthesis of all three. He slowly treaded through the lowly masses of context-dependent survey data, abstracting and categorizing as he went along, relying on previous categorizations for help (see Della Rocco *et al*., 1990), in fact comprising researchers from (Albright *et al.*, 1996). Using these, Ross allows us to see how we can avoid jumping to conclusions in a big unverifiable leap. Instead, we must ascend from the masses of data, and away from them, in small, traceable steps, through multiple intermediate levels of analysis and make it all very explicit and open. Such analysis and synthesis leaves others a trace to follow and critique. It is a consistent reminder to the human factors community (Woods, 1993; Xiao & Vincente, 2000) that if we want to responsibly draw conclusions from our context-specific data (extracted either from the lab or the field), we have to conduct some sort of epistemological analysis. In going from local, particular behaviors to generalizable universal patterns of cognitive strategy, we have to be deliberate, slow and cautious, and above all, explicit. The typical way proposed is to move up from the context-specific details to the concept-dependent generals in a series of successive steps, each more abstract than the previous one; each less in domain terms and more in cognitive terms than the previous one (Xiao & Vicente, 2000). Many refer to this as the process of induction—reasoning from the particular to the general (e.g. Beyer & Holtzblatt, 1998).

Indeed, epistemological analysis, the careful movement from context to concept, is crucial for migrating to designable futures; for making a link between ethnographic analysis and design guidance. This is exactly what study III encourages us to do. For example, context-specific controller activities such as "entering a pilot report; composing a flight plan amendment" reveal a cognitive strategy (Della Rocco *et al*., 1990) at a slightly higher level of analysis: that of the "transformation or translation of information for entry into the system" which, at an even higher level of analysis, could be called "coding", together with other cognitive strategies (Ross, 1995). Part of this coding is symbolic, in that it uses highly condensed code language (red underlinings, black circles) to mark up the flight strip in ways that are meaningful only to the controller him/herself and a few initiated others. Only from there can we make the (then no longer so large) jump to the highest level of abstraction—helping us identify a major cognitive role for the flight strip: the compression of complexity. Unable to keep all the details of what a flight would do stable in the head, the controller compresses this complexity, or amortizes it, as David Kirsh would say, by letting one symbol stand for complex concepts and interrelationships, some even temporal.

Other high level, concept-dependent roles of the flight strip would be the anticipation of dynamics (what comes next) and the support of coordination (e.g. in so-called handovers of flights to other controllers). Note how the language is no longer cast in that of the context-specific details. It is no longer locked into that of a practice intimately connected to the artefact. It is a language that developers can begin to look at as a designable future. Complexity and dynamics, as well as coordination, are critical to what makes air traffic control what it is, including difficult. Whatever developers want to develop, they will have to take into account that controllers use their artefact(s) to help them deal with complexity, to help them anticipate dynamic futures, and to support their coordination with other controllers. This, backed up by its detailed analysis and synthesis, is a useful contour of how to create a designable future. It is all in sharp contrast to Study I which made its jump from context-specifics to one central (and misguided) concept in one large leap, leaving no trace for others to follow or critique, and in contrast to Study II which got stuck in the context-specific trenches, parroting the language of the native. Study III shows us the paradox in contextual inquiry for HCI: Creating designable futures requires extreme sensitivity to context. Yet it asks us to extract the description of people's work *away* from the current context that helps shape it. Otherwise designers will not understand what we are trying to say, and will not know what to do next.

Main Points

- Contextual inquiry is seen by many, especially in CSCW, as one route to human-centered HCI systems. It can reveal what cognitive work means to the people who are doing it, where they are doing it and while they are doing it. If designers tailor the system to how users themselves see their work and their world, then this has to bring designers closer to creating human-centered systems.
- Contextual inquiry is difficult. In order to use contextual inquiry for purposes of requirements capture, we have to make the jump from our description of people's current practice to a design language targeted at the future. We have to untangle what informants do and tell us from what ethnography is and what *it* can tell us.
- Strong contextual inquiry can make this jump. To do so, it has to contain an explicit epistemological analysis. Typically this is a process of induction, moving from context-specific particulars to concept-dependent cognitive descriptions of performance, through multiple cautious steps. The re-use of native categories must be resisted, as must be the making of single leaps from context-specifics to cognitive constructs.
- Methodological imperialism is counterproductive. The contrast between the three studies in this paper shows that if human performance data gathering—however long or however wide—is not backed up by strong analysis, then much goes to waste and designers become misguided as a result. Contextual inquiry is methodologically non-dogmatic. What matters is not so much how you get the data, but what you do with them.
- "Designable futures", and by extension human-centered HCI systems, can result if we succeed in describing people's work in universal cognitive terms that allow designers to start tailoring for the challenges of that work in a proactive way. Designers don't build artefacts or systems so much as they create new ways in which practitioners must handle the challenges associated with their work.

References

Albright, C. A., Truitt, T. R., Barile, A. B., Vortac, O. U., & Manning, C. A. (1996). *How controllers compensate for the lack of flight progress strips* (Final report DOT/FAA/AM-96/5). National Technical Information Service, Virginia.

Anderson, R. J. (1994). Representations and requirements: The value of ethnography in system design. *Human-computer interaction, 9*, 151-182.

Beyer, H., & Holtzblatt, K. (1998). *Contextual design: Defining customer-centered systems*. San Diego, CA: Academic Press.

Della Rocco, P. S., Manning, C. A., & Wing, H. (1990). *Selection of air traffic controllers for automated systems: Applications from current research* (DOT/FAA/AM-90/13). National Technical Information Service, Virginia.

Forsythe, D. E. (1999). 'It's just a matter of common sense': Ethnography as invisible work. *CSCW, 8*, 127-145.

Harper, R. H. R. (2000). The organisation in ethnography: A discussion of ethnographic fieldwork programs in CSCW. *CSCW, 9*, 239-264.

Hollan, J. Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction, 7*(2), 174-196.

Hollnagel, E. (1999). From function allocation to function congruence. In S. W. A. Dekker, & E. Hollnagel (Eds.), *Coping with computers in the cockpit*, pp. 29-54. Aldershot, UK: Ashgate Publishing Co.

Hughes, J. A., Randall, D., & Shapiro, D. (1993). From ethnographic record to system design: Some experiences from the field. *CSCW, 1*, 123-141.

Giddens, A. (1991). *Modernity and self-identity*. San Francisco, CA: Stanford University Press.

Mackay, W. E. (2000). Is paper safer? The role of paper flight strips in air traffic control. *ACM/ Transactions on computer-human interactions, 6*(4), 311-340.

Norros, L., & Klemola, U. M. (1999). Methodological considerations in analysing anaesthetists' habits of action in clinical situations. *Ergonomics, 24*(11), 1521-1530.

Nyce, J. M., and Löwgren, J. (1995). Toward foundational analysis in human-computer interaction. In P. J. Thomas (ed.), *The social and interactional dimensions of human-computer interfaces*. Cambridge, UK: Cambridge University Press.

Plowman, L., Rogers, Y, & Ramage, M. (1995). What are workplace studies for? *ECSCW '95. Proceedings of the Fourth European Conference on Computer Supported Cooperative Work*, Stockholm, Sweden, 10-14 September, 1995. Dordrecht, NL: Kluwer, pp. 309-324.

Ross, G. (1995). *Flight strip survey report.* Canberra, Australia: TAAATS TOI.

Singer, G., & Dekker, S. W. A. (2000). Pilot performance during multiple failures: An empirical study of different warning systems. *Journal of Transportation Human Factors, 2*(1), 63-76.

Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive and healthy computer-based work*. Mahwah, NJ: Lawrence Erlbaum Associates.

Woods, D. D. (1993). Process tracing methods for the study of cognition outside the experimental psychology laboratory. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods*, pp. 228-251. Norwood, NJ: Ablex.

Woods, D. D., Patterson, E. S., Corban, J. M., Watts, J. C. (1996). Bridging the gap between user-centered intentions and actual design practice. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting*, pp. 967-971.

Woods, D. D., & Dekker, S. W. A. (2001). Anticipating the effects of technology change: A new era of dynamics for Human Factors. *Theoretical Issues in Ergonomics Science, 1*(3), 272-282.

Xiao, Y., & Vicente, K. J. (2000). A framework for epistemological analysis in empirical (laboratory and field) studies. *Human Factors, 42*(1), 87-101.