To appear in Proceedings of Safety-critical Systems Symposium 2005

The Effects of Timing and Collaboration on Dependability in the Neonatal Intensive Care Unit

Gordon D. Baxter
Department of Psychology, University of York
York, UK

Juliana Küster Filipe
School of Informatics, University of Edinburgh
Edinburgh, UK

Angela Miguel
Department of Computer Science, University of York
York, UK

Kenneth Tan
Department of Paediatrics, St James's University Hospital,
Leeds, UK

1 Introduction

Computer-based systems are now routinely deployed in many complex dynamic domains, such as aviation and industrial process control. The critical nature of these systems means that their operators rely on them to do the right thing at the right time when called upon. In other words, they are expected to have a high level of what Laprie (1995) defines as dependability. To date dependability research has largely focused on developing techniques for improving the dependability of hardware and software in safety critical applications (e.g., Leveson, 1995). Dependability, however, is a property of the whole socio-technical system: people, computers and context. It is therefore important not only to understand these components, but also how the interactions between them affect dependability.

A wealth of research into human-machine interaction (HMI) has emerged over the last two decades (e.g., see Baecker & Buxton, 1987; and Baecker, Grudin, Buxton, & Greenberg, 1995). One obvious aspect of HMI that affects dependability, is the temporal properties of the interaction. Delays in system response times, for example, can make tasks more complex, and may lead to errors (Johnson & Gray, 1996). Such issues are present in most complex dynamic domains, but particularly in real-time systems, including medicine (e.g., Combi & Shuhar, 1997).

This paper considers how timing issues in HMI affect dependability in one specialised branch of medicine: neonatal intensive care. A case study was carried out in the Neonatal Intensive Care Unit (NICU) at St James' University Hospital (SJUH), Leeds. An expert system, FLORENCE (Fuzzy LOgic for REspiratory Neonatal Care Expert), is being developed at SJUH (Tan et al., 2003) to help less

experienced staff make decisions about changes to the ventilators that are regularly used in treating premature babies.

One of the goals of the study was to identify any timing issues involved in the current socio-technical system that implements the practice of neonatal intensive care. More particularly, the aim was to identify and analyse those aspects of timing that affect how staff interact with the equipment in the NICU which places requirements on the design and use of FLORENCE. Once FLORENCE is in place, the dependability of the new system (including FLORENCE) should be at least equal to the dependability of the system without FLORENCE.

Section 2 of the paper provides an overview of the NICU environment at SJUH and an overview of the case study. Sections 3 and 4 highlight and analyse the timing and collaboration issues identified by the study and how they contribute to the dependability of the system of patient care. Section 5 examines how the introduction of FLORENCE could affect the timing and collaboration aspects of dependability. Section 6 briefly discusses the completeness of the case study methods in identifying timing and collaboration issues and considers the available alternatives. Finally, Section 7 summarises the work and suggests how it could be extended in the future to evaluate the impact of FLORENCE.

2 The Neonatal Intensive Care Unit

When a premature baby arrives in the NICU it is often placed on mechanical ventilation to help it deal with respiratory ailments, particularly Respiratory Distress Syndrome (RDS). This is a self-regulating disease, caused by the lungs not having developed sufficiently to produce the levels of surfactant required to facilitate gaseous exchange during respiration (e.g., Rennie & Roberton, 2002). The main aim during the period of ventilation is to stabilise the baby, such that its blood gas and pH levels remain within some predefined range. These parameters which are continuously monitored using a Neotrend multi-parameter intra-arterial sensor (Philips, 2001), cannot be directly controlled. They are affected by the baby's respiration, however, which is controlled using a Babylog 8000+ ventilator (Bartholomew, Newell, Dear, & Brownlee, 1994).

Much of the front line care of the babies is performed by nursing staff and junior doctors (Senior House Officers, SHOs). The SHOs normally only work in the NICU for six months as part of their job rotation. In general, the SHOs perform most of the interventions on the ventilator in acute situations where the baby has RDS, calling on more experienced members of staff as appropriate. One of the goals of FLORENCE is to empower the front line carers by helping them to more rapidly learn to make correct decisions about which interventions to make.

A Cognitive Task Analysis of the work in the NICU was performed in three stages (for full details of the methods see Baxter, Monk, Tan, Dear, & Newell, Submitted). First, domain and context familiarisation was carried out. Second, the Critical Decision Method (CDM; Klein, Calderwood, & MacGregor, 1989) was used to analyse decisions surrounding use of the ventilator. Third, naturalistic observation of the use of the ventilator in situ was carried out. These stages are briefly summarised below.

The first stage involved bootstrapping into the domain. In order to understand the language of the domain and the physical and social context in which FLORENCE was to be deployed, a lightweight version of the rich picture method was used (Monk, 1998). Eight members of staff at the unit were interviewed (including one administrator) to identify their roles and responsibilities.

In the second stage the CDM was used to analyse the decision making processes involved in making interventions using the ventilator. Semi-structured interviews were carried out with eight members of staff: four front line carers (nurses and SHOs) and four experts (registrars and consultants). The interviewees were asked to recall incidents where they had been involved in making decisions about changes to the ventilator settings.

The final stage involved a period of observation of work at the unit, focusing solely on a single baby. An earlier neonatal care study which used observation periods of between one and two hours, found that interactions with computerised equipment were fairly infrequent (Alberdi et al., 2001). Their observations were carried out at various times of the day over several months, and related to the unit as a whole, rather than just the NICU. Alberdi et al. were concerned mainly with the way that people used computerised monitoring equipment in neonatal care per se, whereas the goals of this study are more tightly constrained. Here, the main aim was to identify timing issues around use of the ventilator in dealing with RDS, with a view to informing the development of FLORENCE. Since RDS is a self-regulating disease which only lasts a matter of a few days, the baby normally recovers within about a week. Any interventions involving changes to ventilator settings will thus tend to be concentrated within that period. It was therefore decided to use observation sessions lasting two hours on each of the days when a baby was connected to the ventilator.

The findings of the case study are summarised in Table 1 as implications for the design and use of FLORENCE. In addition to these findings, several timing and collaboration issues were also identified which relate to the wider socio-technical system of the NICU. These issues are described in more detail below, where their implications for dependability are also considered.

3 Timing

In health care, the introduction of new technology often disrupts the socio-technical system in which the technology is embedded. The way that information is distributed, and the tasks that are performed are adversely affected, which has a knock-on effect on the relationships between health care professionals and other staff (Berg, 2001). It is therefore important to understand the timing and collaboration issues that exist before the new technology is introduced, so that the impact of the new technology can be properly evaluated. The main concern here is to understand how these issues influence the dependability of the socio-technical system.

Staff need to be aware of any contingencies (such as ignoring anomalous data) before implementing FLORENCE's suggestions.
FLORENCE should prompt staff to follow the DOPE mnemonic.
The FLORENCE audible alarm needs to be distinctive.
Staff need to be trained how to respond to a FLORENCE alarm.
The size of the text used to display FLORENCE's advice needs to legible when staff are stood at the ventilator.
The wording of FLORENCE's suggestions needs to be clear and unambiguous (e.g., PIP Up by 2 to 16)
Space needs to be made available for the PC running FLORENCE, and a mouse (unless a touch screen is used).
Consideration needs to be given to what data from FLORENCE needs to be included with existing paper records.
It must be possible to print data from FLORENCE for inclusion with the other patient records.
There needs to be a power socket available for the PC running FLORENCE.
The limitations of FLORENCE need to be made explicit to staff.
FLORENCE should check the current ventilator settings before deciding what changes are required.
Staff need to be made aware of the potential data redundancy problem, because FLORENCE will display trends of data that are available on the ventilator and Neotrend.
FLORENCE should be able to explain its decisions on request.
Staff should be able to override FLORENCE's suggestions, as long as they can supply a reason for doing so.
FLORENCE should attempt to avoid generating intermittent alarms.

Table 1. Implications for the design and deployment of FLORENCE identified by the case study.

The control of a baby's blood gases has some similarities to industrial process control. The basic control loop of perceiving, deciding and acting, for example, is common to both. The situation in the NICU is, in several respects, more complicated, however. Whereas continuous data sampling has been routine in process control for many years, it is only through the recent introduction of the Neotrend that reliable blood gas data can be collected continuously. Staff in the NICU have to effect control by monitoring the babies condition and, where appropriate, adjusting the ventilator settings and administering the necessary drugs.

The situation in the NICU is also somewhat more complex because the problems faced by an individual baby are partly determined by its gestation period at birth, and its birth weight. Generally, the closer a baby is to the normal term of a pregnancy (40 weeks) at birth the less likely it is to suffer from RDS, because its lungs will be more fully developed. These factors affect how long it takes to treat an individual baby and stabilise its condition.

After the ventilator has initially been configured, it will normally only be changed in response to acute situations. The decision to change the ventilator

settings is normally made after discussions between a team of people. This is followed by the physical action of changing the settings. There are obvious timing and collaboration issues involved here.

3.1 Timing Issues

Examples of the timing issues that were explicitly described by clinical staff during interviews are listed in Table 2. It should be noted that the results of the CDM emphasised that the overriding concern of the NICU staff is the clinical outcome, rather than the amount of time available to stabilise the baby's condition.

3.2 Timing and Dependability

In dynamic complex domains such as medicine, time is an inherent determinant of system dependability: if the system cannot produce a safe outcome in a timely manner, then it cannot be regarded as dependable. There is a trade-off that needs to be made in the NICU between time and the quality of the decision making (and the associated action). So, for example, if a better decision can be made by waiting for a short time to get access to some more data, this may be preferable to starting on one course of action and then having to radically change it shortly afterwards.

The areas of dependability in the NICU that are most affected by timing issues are reliability and safety. If the system does not always produce appropriate responses in a timely manner, this could affect the well being of the baby.

The system has been developed on the basis of experience and best practice. In some cases timing issues have been introduced to impose a work structure that increases the system dependability. The use of deliberate delays before responding to alarms, for example, is a response to the inherently noisy data that is generated by the system (Miksch, Seyfang, Horn, & Popow, 1999). When a baby moves, for example, the heart rate increases, which causes the heart rate alarm to sound. In such situations the alarm may only sound a few times at most, so staff tend to wait to see if the alarm continues before responding. This allows any transient alarms to clear, but as a precautionary measure staff will also glance at the baby to check for any immediately apparent problems.

The pacing of work in the unit is the result of a trade-off between the temporal validity of data and the need to allow the babies to rest as much as possible. There are two main aspects of pacing. The first is the recording of hourly observations, which is implemented in a flexible manner. The observations for a particular baby can be brought forward or delayed as necessary to accommodate any changes in staff workload without adversely affecting the baby. The second is when the baby is not connected to a Neotrend. In this case a regime is put in place to take manual blood gases at regular intervals (3-4 hours), which involves pricking the baby's heel to collect the blood. This causes the baby some discomfort which partly explains the long interval between successive manual blood gas tests. Temporal validity is also an issue that affects the accuracy of the Neotrend readings which decay over time. When the data starts to decay, the Neotrend has to be recalibrated.

Category	Examples
Sequence	There is a sequence of systematic checks performed during diagnosis: DOPE (displacement of the ETT, obstruction of the ETT, pneumothorax, equipment failure). Manually bag the baby before carrying out suction. After changing the ventilator settings, take a manual blood gas to confirm the effect of the changes (after 15-30 min delay). Check the baby's status; decide if any action is necessary; take the appropriate action. Check the baby's Airway, Breathing and Circulation (ABC) in that order.
Delayed feedback	It takes 2-3 minutes before the rise in the PO2 brought about by the changes becomes apparent. The full rise takes about 20 minutes. The effects of the changes are not quantitatively assessed until a manual blood gas is taken. X-rays and manual blood gases have to be processed before the results are available.
Delays	A deliberate delay was introduced in the incident to allow the implemented changes to take effect and to let the consultant assess the changes that were made. Configuring the high frequency oscillator ventilator (HFOV) takes 20 minutes. This had to be considered when deciding whether to switch the baby onto HFOV. Staff deliberately wait before responding to alarms, to check whether the alarm is real.
Trends	Staff start looking for trends in the blood gases that show an improving situation after about 2-3 minutes. The Neotrend display shows the effects of the changes to the PIP and the TI as changes in PO2 and PCO2. Indications that the blood gases in particular are heading in the right direction.
Temporal validity	Manual blood gas data is only valid for about 20-30 minutes. Chest X-rays are only valid for a limited period of time. Manual blood gas data decays over time. The Neotrend is periodically recalibrated to maintain data accuracy.
Deadlines	If the oxygen level is critically low, then something has to be done in very few minutes, with an improvement in condition required in 10-15 minutes. Self imposed deadlines for detecting improvement in baby's condition.
Pacing	Observations of the babies condition are taken and recorded on an hourly basis. Where a Neotrend is not being used, blood gases are sampled regularly (normally every three or four hours).
Other	There are temporal aspects to the way that the ventilator operates, in that the inspiration and expiration time can be controlled.

Table 2. Summary of timing issues.

There are emergent timing loops in the work structure which reflect the defence in depth mechanism that allows cases to be referred to more expert staff in a timely manner. The front line carers tend to handle the smallest and simplest problems, which also take the shortest time. If a problem is more complex a registrar is typically called in to assist, and solving the problem will take longer. For the most complicated problems a consultant may also subsequently be called in, and these problems generally take the longest time to solve. Although these control loops are all based on competence, they are also characterised by their execution times.

The use of sequences is another example of an imposed feature. Although the DOPE mnemonic does not require a temporal ordering of the checks, the intention is to make the system safer by making sure that some of the common alternative causes of acute situations are not overlooked before changes are made to the ventilator settings. The ABC mnemonic sequence, however, does have a strict temporal ordering, based on getting the system into a safe state at the earliest opportunity.

Whenever staff were asked about deadlines and time pressures, they invariably replied with a range of values—a few minutes, seconds to minutes, and so on—rather than a hard deadline. It could be argued that this use of less tightly specified deadlines is an attempt to make the system safer and more reliable by not imposing unnecessarily strict time pressures on staff who are already working in a stressful environment. Furthermore, where deadlines were identified these tended to vary between babies, because they are largely determined by the individual baby's physiology.

The remainder of the timing issues can be considered as inherent properties of the components of the system, and hence need to be noted when assessing the system's dependability. There are natural delays between actions and the results of those actions being observable: the settle times of the equipment and the baby's physiology contribute to the delayed feedback and temporal validity of the data. When the ventilator settings are changed, for example, some effect can be observed shortly afterwards, but it takes 20-30 minutes for the full effects to become apparent. Similarly, X-rays take 20-30 minutes to be developed. Whilst these are aspects of the system that cannot be directly changed by staff they can be accounted for because they have known, relatively short limits. It is much more difficult, however, to gauge the full effect of some of the interventions when they are made. Where complications arise, due to oxygen starvation for example, the full extent of any problems may not become known until several months or even years afterwards.

The Neotrend displays trends of continuous blood gas data (partial oxygen and partial carbon dioxide pressures, and pH), which means that staff can be more responsive after changing the ventilator settings. This allows them to check that the changes are affecting the blood gases in the right direction at an appropriate rate and then respond accordingly.

4 Collaboration

In addition to the timing aspects of the interaction between the humans (NICU staff) and the technological equipment, the system also depends on the interactions between staff. Examples of the different collaboration issues that were identified are

summarised in Table 3. These issues of human collaboration are considered in more detail below.

4.1 Collaboration Issues

The unit employs about 70 staff with varying levels of expertise. Collaboration between staff is essential to the successful functioning of the unit. The nature of this collaboration and how it is supported and facilitated by the infrastructure of the unit directly influences the system dependability.

Category	Examples
Organisational structure	Communication hierarchy. Decision making hierarchy in which nurses and SHOs make initial decisions, but can refer problems to registrars who can refer them to consultants.
Formal verbal communication	The ward round acts as the formal shift handover between the night shift and the day shift.
Informal verbal communication	Separate informal shift handovers for the nurses and the SHOs.
Formal non-verbal communication	Records are used extensively in the NICU by all staff.
Informal non-verbal communication	Overseeing Overhearing Changes to equipment settings

Table 3. Summary of collaboration issues.

4.2 Collaboration and Dependability

The organisation of work in the NICU makes it important that the staff work as a team to deliver an effective service of patient care. The corollary of this is that the need for sharing information and knowledge is critical to the system's dependability.

The main form of collaboration is communication between staff. Much of the information flow in health care takes place in the clinical communication space (Coiera, 2000). The unit operates a shift system, and communication takes place during and across shifts. Communication is also one of the major means through which other aspects of collaboration are achieved (Bardram, 1998). It is used to achieve coordination, cooperation, negotiation, planning, decision making and generally for sharing information. Decisions and actions concerning the ventilator settings are based on communication between staff, and the data values displayed by the monitoring equipment in the NICU.

Senior staff have several roles including clinical care, management and staff training. There is also some overlapping of roles. Some senior nursing staff, for example, have specialist training that allows them to perform more specialised medical tasks, like the SHOs, and the registrars may perform medical tasks like the SHOs, and make managerial decisions like the consultants.

In addition to the doctors and nurses several other staff play important supporting roles in the unit. These include technical support, ward clerks who collate various information sources generated by the unit, and a pharmacist.

The overlap between the roles and responsibilities of staff contributes to the reliability and availability of the system. If there is a shortage of SHOs due to illness, for example, their tasks can still be performed by the registrars, making the system resilient to single point failures.

The safety and reliability of the system are also helped by the hierarchical structure of decision making. This provides a defence in depth mechanism, in that if an SHO is having problems that they cannot solve with a particular case, they can refer it upwards to the registrars who can, if necessary, refer it to the consultants. The referral process is determined by the individual's level of expertise and the complexity of the case.

There are potential problems if knowledge is distributed too widely, at least in geographical terms. The staffing levels at the unit are coordinated in such a way that there is always an appropriate pool of knowledge available to deal with most situations. For the most difficult cases, the consultants are always available by telephone.

The sharing of information is used to generate the bigger picture for each case and to facilitate the coordination of staff. This is achieved through a combination of formal verbal communication—mainly the ward round—and informal verbal communication, such as the other shift handover meetings. These meetings allow work to be planned and co-ordinated for the next time period, and hence facilitate co-operation by giving staff an awareness of what others will be concerned with during the same period. The need for a shared up-to-date awareness of the situation is essential for dependability and provides a basis for the team to perform several tasks concurrently.

The main shift handover occurs during the ward round each morning, when the night shift hands over to the day shift. This formal handover is used to establish a care plan for the next 24 hours for each of the babies. The ward round is normally led by one of the consultants, and is attended by registrars, SHOs and nursing staff. Everyone is given the chance to contribute as individual cases are discussed and decisions explained at length. In this way the ward round also provides a means of in-service training.

Whilst the formal verbal communication provides staff with the key information needed to perform their work, it is the informal verbal communication that allows the system to continue to function effectively in the face of dynamically changing situations. Three examples of informal verbal communication are described below.

In addition to the ward round, the nurses, registrars and SHOs each have their own less formal staff handover meetings. These allow staff to share more detailed information using their own terms and decide how to achieve the care plans laid out during the ward round.

During the CDM several staff reported using explicit verbalisation by talking aloud when diagnosing a particular case. Verbalisation can save time because all of the staff who are present are simultaneously kept in the loop and hence can act

immediately. Staff are also kept in the loop by overhearing the verbal communication of others.

Verbalisation also gives other staff some insight into the decision making processes for particular cases, and allows them to provide input to decisions. Furthermore, it provides a verbal audit trail which can be useful during subsequent discussions, enabling staff to reflectively learn from the handling of the case.

Whilst working, staff frequently talk to each other. Most explicit verbal communication is face-to-face because the people that could be required to perform a task are usually present on the unit. The main exception is when the on-call doctor has to be contacted by telephone or pager. Informal verbal communication is important for coping with changes in plans, either by sharing new information or negotiating new plans; for requesting help, and for discussing the specifics of how to achieve planned goals.

The consultants particularly rely on verbal communication with the front line carers to bring them up to date with the evolution of events. In acute situations this is the quickest way for them to get a current assessment of the situation. They utilise the communicated information together with the available data from the monitoring equipment, such as trends of the blood gases, and patient charts to help guide their decision making. There is lots of informal communication between the registrars and the consultants.

Both formal and informal verbal communication rely on the supporting infrastructure of the unit, particularly the patient records. Much of the communication in the NICU involves references to records. The records allow information to be shared between members of staff without them having to speak to one another. This can save time and reduce the amount of information that has to be remembered. Records are particularly useful for passing information between people who are present in one place at different times, as happens across shifts. This contributes to the reliability of the system because it supports the continuity of patient care. Over longer periods the records also provide a form of backup, which contributes to the integrity of the system.

Non-verbal communication is used to structure the work and record information for audit purposes, and provides a back-up of information. A written history of conditions and treatments is provided by the patient records, which can be used to resolve problems or ambiguities for individual cases; active problems are recorded using problem sheets (currently only for the doctors). The patient records also provide some level of data redundancy in that they are a written back up of staff knowledge about particular cases. The doctor's notes, for example, record relevant decisions made during the ward round, and details of subsequent events for the individual cases. Other documents contain information relating to the required treatment of the patient and any test results. Plans are not usually written down because they frequently have to be revised dynamically in response to changing situations.

There are several other informal ways in which non-verbal communication occurs. These alternatives to speech support verbal communication and provide faster ways of passing information when time is at a premium. Informal notes also allow information to be shared between people who cannot meet in person.

The main form of informal non-verbal communication used in the unit is passive communication. This occurs when one member of staff simply watches what another is doing, and when someone notes the changes that have been made to equipment settings and records and uses these to infer the actions of others. The latter can be triggered when somebody notices that the monitoring equipment is showing a change in the trend of the data values that are being recorded, for example.

The concept of dependability has recently been extended to take appropriate account of all aspects of socio-technical systems, such as patient care in the NICU (Dewsbury, Sommerville, Clarke, & Rouncefield, 2003). One of the aspects that has been added is learnability, and this is certainly applicable to the NICU. The SHOs only work in the NICU for six months on rotation as part of their general training to be a doctor. During that period they are expected to learn about patient care in the NICU.

The ward round provides a relatively formal way of learning about the system of patient care. The use of verbalisation by senior staff also contributes to learnability by helping the SHOs (and others) to understand about decision making task performance in the NICU. Even where verbalisation is not used, simply watching others provides another useful way of learning about the system.

The patient records also contribute to the learnability of the system by helping to structure verbal communication. The most obvious example occurs during the ward round where the daily update book is used as the source for discussions about each of the cases.

5 The Impact of FLORENCE on Dependability in the NICU

The way in which timing and collaboration features help to make the NICU dependable have been identified above. When FLORENCE is introduced, it should make the NICU even more dependable by helping front line carers to get babies into a more stable (safer) condition more rapidly. It should help to overcome the tendency of front line carers to be more conservative than experts when deciding on the magnitude of changes to the ventilator settings: the front line carer make smaller changes which means that they subsequently have to repeat the decision making and action processes.

What is not clear, however, is the impact of FLORENCE on the other aspects of work in the NICU, which could have a knock-on effect on dependability. Below, consideration is given to how FLORENCE could affect timing and collaboration in the NICU, together with how the impact can be assessed and appropriately managed.

The inherent timing issues identified in section 3 will not be affected by FLORENCE, because they depend on the properties of the equipment and the baby's physiology. FLORENCE could influence some of the other timing aspects, however. As noted above, the data in the NICU is inherently noisy; FLORENCE could exacerbate this problem if it generates more artefactual (false) alarms. The system of patient care has already developed a way to deal with this problem: staff generally do not respond immediately to alarms before checking the nature of the alarm.

Furthermore FLORENCE has been designed to remove some of the noise from the data used in calculating whether changes to the ventilator settings are required. These two defence mechanisms mean that FLORENCE should improve dependability by making the system safer and more reliable, because decisions will be based on better quality data.

There is a potential impact on learnability of the system. If it is fairly evident to staff what decisions FLORENCE is recommending and why, they may not verbalise the reasoning behind the decision to the same extent. This obviously reduces the scope for learning from other staff. This has been compensated for in FLORENCE by providing a facility to explain why it has made a particular decision. The system will become less safe, however, if staff always blindly accept FLORENCE's suggestions. The effects on learnability can be mitigated during the induction period for the SHOs when it needs to be clearly explained that FLORENCE is an *advisory* system, and that its decisions should not always be accepted without question.

By making learning faster, FLORENCE could reduce collaboration. This is because front line carers may become more self-reliant (if FLORENCE is trusted by staff) and able to respond to situations without needing to interact with others, particularly the more experienced staff. The corollary of this is that there will be fewer opportunities to learn from the more expert members of staff. This will affect the learnability of some parts of the system, and may also affect the safety and reliability if the front line carers do not learn about those subtler aspects of the system that are not encompassed by FLORENCE. These effects can also be addressed during the SHOs induction period. Staff should be encouraged to try and work out what changes they would make, and then explain the reasons why. Their suggestions could be compared to FLORENCE and any differences discussed with other staff.

Similarly, a reliance on FLORENCE could lead to fewer gatherings around the baby's cot to discuss interventions. Other nurses would therefore have fewer opportunities for picking up information through overhearing, hence reducing their overall awareness of what is happening in the NICU. The net effect could be a reduction in the system's safety and reliability if it means that it takes staff longer to get up to speed when they get called in to help with cases where they have not previously been directly involved. More senior staff in particular should be encouraged to make sure that they continue to discuss interventions and to encourage the SHOs to do so to, rather than just relying on FLORENCE.

There are two key aspects to ensuring that FLORENCE does not lead to a less safe system and, more generally, make it less dependable. The first lies in the education of the front line carers. During the SHOs induction training, the role and purpose of FLORENCE needs to be very clearly explained. The second is the monitoring of the use of FLORENCE. As part of the checking of the progress of the SHOs, senior staff will need to monitor the usage of FLORENCE to make sure that front line carers are not becoming too trusting or too reliant on FLORENCE. This will necessarily include checking how much SHOs are learning from communicating and collaborating with other staff. The process of monitoring the use of FLORENCE will obviously be an ongoing one.

6 General Discussion

The case study has highlighted several timing and collaboration issues that affect system dependability in the NICU. The way in which FLORENCE can affect the dependability has also been considered with respect to these issues. There may be some other timing and collaboration issues, however, which have not been identified by the methods used in the case study. It is therefore worth considering what other methods could be used to uncover those issues.

Timing issues in HMI are something that needs to be observed in situ. Routine timing issues, such as periodicity of work can be identified by observation. In acute situations, however, which is where most of the HMI takes place, it is more difficult to identify the precise nature of the timing issues. The CDM can be used to identify timing issues at a general level of abstraction, and in terms of the order of occurrence of events and actions. It is much more difficult to pin down the absolute times at which things happened in the recalled incidents, however. The main problem is that these events are unpredictable in terms of when they will occur, and how long they will last, since they are largely determined by the baby's physiology.

Detailed timing information can often be captured using video recording. In the NICU, however, where the occurrence of interesting events cannot be reliably predicted this would require video recordings over extended periods (possibly days). The NICU is a highly stressful situation for all concerned—babies, parents and staff—which is one of the reasons why the local ethics committee would not sanction the use of an intrusive method like video recording over long periods of time.

There is therefore a problem of how to get at the detailed timing and collaboration issues whilst avoiding potential problems of post hoc rationalisation about those issues. One obvious solution is to shadow the staff as they go about their daily work. This is a time and labour intensive approach and requires close co-operation with those people being shadowed. Perhaps the best described example of this approach is Hutchins (1995), who studied ship navigation by shadowing the operations on the bridge of a ship on various occasions over several months. This approach has also been successfully used to qualitatively explore communication patterns in the field of medicine by Coiera and Toombs (1998).

It is not clear whether all the collaboration issues were identified when using the CDM. This could be due to the method, but may also be attributed to the fact that during the incident being recalled there was so much happening that the interviewee could not physically track what everyone was doing, or whether collaboration was actually reduced. Interviewees also did not produce any detailed timing information associated with the coordination and co-operation of activities. The interviews to build rich pictures of roles, responsibilities and artefacts for non-verbal communication provide more information about collaborative work. They do not, however, provide information about the interaction between people and equipment that goes on at critical times.

7 Summary

Dependability is inherent in the NICU system. Contributions to the dependability of the socio-technical system are made by the humans (the NICU staff), the technology (ventilator, Neotrend and so on) and the context (social and physical). This study has focused more on those aspects associated with the humans, particularly the timing aspects of the human-machine interaction, and collaboration between the staff.

The study identified several examples of timing issues and collaboration issues that contribute to the dependability of the system before the introduction of FLORENCE. It is important that the system after the introduction of FLORENCE should be at least as dependable as it was before. Although there are often problems when new technology is introduced into clinical settings by outsiders (Coiera, 1999), this should not be the case in the NICU because it was the consultants in the unit who identified the need for FLORENCE, and they have been closely involved in its development. Furthermore they have the experiences of a previous expert system, ESNIC (Snowden, Brownlee, & Dear, 1997) to build on, where the issue of having to manually transfer the data readings from the equipment into ESNIC made it unacceptable to staff. The consultants took this into account in their requirement that FLORENCE had to be both clinically useful and acceptable to the staff for it to be successful.

Whether FLORENCE will improve the dependability of the system is an empirical question, albeit a difficult one due to the particular constraints of the NICU. SHOs are on rotation in the neonatal unit for only six months . This, coupled with the fact that the set of cases of RDS in the NICU changes considerably over time, makes it hard to directly compare the pre- and post-FLORENCE systems. The different rates of development of the babies also makes it hard to make direct comparisons across cases that are matched for gestation and birth weight at the start of their stay in the NICU for the pre- and post-FLORENCE systems.

Some of the ways in which FLORENCE could affect the dependability of the system have been identified here. In particular, the possible ways in which timing issues and collaboration could be affected have been highlighted. The next step is to return to the NICU after FLORENCE has been in place for some time to evaluate how the identified issues have really been affected.

Acknowledgements

The authors would like to thank the staff at SJUH Neonatal Unit for their helpful advice, assistance and co-operation during the study. This work was funded by the EPSRC (grant number GR/N13999) as part of the Dependability Interdisciplinary Research Collaboration (DIRC); the development of FLORENCE was funded by the SPARKS (Sport Aiding medical Research for KidS) charity.

8 References

Alberdi, E., Becher, J.-C., Gilhooly, K., Hunter, J., Logie, R., Lyon, A., McIntosh, N., & Reiss, J. (2001). Expertise and the interpretation of computerised physiological data: Implications for the design of medical decision support in

neonatal intensive care. *International Journal of Human-Computer Systems, 55*(3), 191-216.
Baecker, R. M., & Buxton, W. A. S. (Eds.). (1987). *Readings in human-computer interaction: A multidisciplinary approach.* Palo Alto, CA: Morgan Kaufmann.
Baecker, R. M., Grudin, J., Buxton, W. A. S., & Greenberg, S. (Eds.). (1995). *Readings in human-computer interaction: Toward the year 2000*. Los Altos, CA: Morgan Kaufmann.
Bardram, J. E. (1998). Designing for the dynamics of cooperative work activities, *Proceedings of the 1998 ACM conference on computer supported cooperative work* (pp. 89-98). Seattle, WA: ACM Press.
Bartholomew, K., Newell, S., Dear, P., & Brownlee, K. (1994). *Babylog 8000 - Flow wave and volume monitoring* . Lübeck, Germany: Drägerwerk AG.
Baxter, G. D., Monk, A. F., Tan, K., Dear, P. R. F., & Newell, S. J. (Submitted). Using cognitive task analysis to identify requirements for the successful deployment of decision support systems in neonatal intensive care. *Manuscript submitted for publication*.
Berg, M. (2001). Implementing information systems in health care organizations: myths and challenges. *International Journal of Medical Informatics, 64*, 143-156.
Coiera, E. (1999). The impact of culture on technology: How do we create a clinical culture of innovation? *Medical Journal of Australia, 171*, 508-9.
Coiera, E., & Toombs, V. (1998). Communication behaviours in a hospital setting: an observational study. *BMJ, 316*, 673-676.
Combi, C., & Shuhar, Y. (1997). Temporal reasoning and temporal data maintenance in medicine: Issues and challenges. *Computers in Biology and Medicine, 27*(5), 353-368.
Dewsbury, G., Sommerville, I., Clarke, K., & Rouncefield, M. (2003). A dependability model for domestic systems, *Proceedings of 22nd International Conference on Computer Safety, Reliability and Security (SAFECOMP'03)* . Heidelberg, Germany: Springer-Verlag.
Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.
Johnson, C., & Gray, P. (1996). Temporal aspects of usability. *SigCHI Bulletin, 28*(2), 32-33.
Klein, G., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on System, Man, and Cybernetics, 19*, 462-472.
Laprie, J.-C. (1995). *Dependable computing: Concepts, limits, challenges.* Paper presented at the 25th IEEE International symposium on fault-tolerant computing, Pasadena, CA.
Leveson, N. (1995). *Safeware: System safety and computers*. Reading, MA: Addison-Wesley.
Miksch, S., Seyfang, A., Horn, W., & Popow, C. (1999). Abstracting steady qualitative descriptions over time from noisy, high-frequency data. In W. Horn, Y. Shuhar, G. Lindberg, S. Andreassen, & J. Wyatt (Eds.), *Artificial Intelligence in Medicine. Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making, AIMDM'99, Aalborg, Denmark, June 1999, Proceedings. Lecture Notes in Computer Science 1620.* (pp. 281-290). Berlin: Springer.

Monk, A. F. (1998). Lightweight techniques to encourage innovative user interface design. In L. Wood (Ed.), *User interface design: bridging the gap between user requirements and design* (pp. 109-129). Boca Raton: CRC Press.

Philips. (2001). *Trendcare continuous blood gas monitoring system: Technical Overview* . Böblingen, Germany: Philips Medical Systems.

Rennie, J. M., & Roberton, N. R. C. (2002). *Neonatal Intensive Care*. (4th ed.). London: Arnold.

Snowden, S., Brownlee, K. G., & Dear, P. R. (1997). An expert system to assist neonatal intensive care. *Journal of medical engineering technology, 21*(2), 67-73.

Tan, K., Baxter, G., Brownlee, K. G., Newell, S. J., Dear, P. R. F., & Smye, S. (2003). Fuzzy logic expert system for ventilation of the newborn infant. *Archives of Disease in Childhood, 88*(Supplement 1), A47.