The ESA/NASA SOHO Mission Interruption:
Using the STAMP Accident Analysis Technique for a Software Related 'Mishap'

Chris Johnson (a), C.M. Holloway (b)

(a) Department of Computing Science, University of Glasgow, Scotland.
johnson@dcs.gla.ac.uk

(b) NASA Langley Research Center, MS 130 / 100 NASA Road, Hampton, VA 23681-2199, USA
c.m.holloway@larc.nasa.gov

Abstract: Mishap investigations provide important information about adverse events and are intended to help avoid any recurrence of previous failures. However, the complexity of many safety critical systems poses new challenges for mishap analysis. Similarly, the recognition that many failures have complex, systemic causes has helped to widen the scope of many mishap investigations. A new generation of mishap analysis techniques has been proposed to help investigators address these problems. For instance, Leveson has recently developed the Systems Theory Accident Modelling and Process (STAMP) approach to address some of the weaknesses associated with previous 'chain of event' approaches that can miss the systemic causes of adverse events. There are relatively few examples of the STAMP approach. This paper, therefore, presents the results obtained when two analysts performed an independent application of this technique to analyse the causes, including software problems, which led to the mission interruption of the joint European Space Agency (ESA) and National Aeronautics and Space Administration (NASA) Solar and Heliocentric Observatory (SOHO).

Keywords: STAMP, Causation, SOHO, Accident Investigation, Mishap Investigation, Software Mishaps.

Introduction

Many international standards now require that mishap-reporting systems be integrated into safety management schemes. For example, IEC 61508 is widely used as a standard for the development of safety-critical applications that incorporate computer systems. This includes the recommendation that manufacturers should: "…implement procedures which ensure that hazardous Mishaps (or Mishaps with potential to create hazards) are analysed, and that recommendations are made to minimise the probability of a repeat occurrence." (IEC, paragraph 6.2.1). NASA has responded to these initiatives by developing sophisticated support for mishap reporting. Each individual sites provides faculties for reporting local safety concerns. There are also more centralised applications, such as NASA's Safety Reporting System and a lessons learned system.

More serious mishaps trigger the formation of investigation boards. These are guided by the requirements of NASA Procedures and Guidance document 8621.1. This provides information on appropriate analytical techniques. It also enumerates the key documents that must be produced during the investigation process. Such guidance helps to ensure a consistent approach to the analysis of different mishaps. This is important given the diverse range of safety-critical applications that are operated by NASA. The techniques that are recommended in NPG 8621.1 address a number of common problems in the investigation of adverse events. For instance, they include advice on the reconstruction of the events leading to a mishap using technique that are similar to timelines. This reconstruction task is important because it can be difficult to build up a complete picture of what went on during an incident. Evidence is often missing and eyewitnesses frequently contradict each other. These reconstruction techniques are often referred to as 'chain of events' models.

NPG 8621.1 also recommends methods for performing a causal analysis once a reconstruction has been completed. Many of these techniques date back to pioneering work conducted by the US Department of Energy and the National Transportation Safety Board during the 1970's and 1980's. Approaches such as Multilinear Event Sequencing and Events and Causal Factors Charting first develop the 'chain of events' reconstructions mentioned above and then provide techniques for analysing them to distinguish root causes from contributory factors (Johnson, 2003). The approaches recommended in NPG 8621.1 have been widely applied to support the investigation of incidents in industries ranging from oil production and

transportation through to healthcare and pharmaceutical manufacture. However, a number of caveats and criticisms have recently been raised about them. They were initially intended to analyse the types of incidents that were common when they were developed. It can, therefore, be argued that they are ill suited to analyse the types of tightly integrated computer-controlled systems that typify many current safety-critical applications (Leveson, 2002). It has also been argued that these early techniques do not reflect the current focus on the systemic causes of failure. In consequence, they can be used to identify causal sequences without necessarily helping investigators to identify the mass of different contextual, causal factors that create the preconditions for failure (Reason, 1997).

A number of techniques have recently been proposed to address some of these limitations. For instance, Ladkin and Loer's (1998) Why-Because Analysis provides the benefits of formal reasoning about the causes of adverse events and near misses. The intention is that mathematical proof techniques will increase our confidence that any analysis is justified in terms of the evidence that is available during an incident investigation. However, this approach still relies upon an initial reconstruction that develops a 'chain of events', similar to those embedded within the previous generation of analytical techniques. This reconstruction is used to enable domain experts to validate the model of adverse events before formal reasoning can begin. Leveson (2002) has recently presented a strong criticism of these 'chain of events' techniques. Her objections focus on the manner in which such models 'can easily miss subtle and complex couplings and interactions among failure events and omit entirely accidents involving no component failure'. Event-chain models cannot easily be used to analyse systemic failures including 'structural deficiencies in the organisation, the safety culture in the industry, and management deficiencies'. Leveson, therefore, proposes a control-based approach called the Systems Theory Accident Modelling and Process (STAMP). A control-model is constructed once investigators have identified the sequence of events leading to an incident. The intention is not to completely eliminate event reconstruction but to introduce additional stages of analysis that focus more on the systemic causes of adverse events. The remainder of this paper focuses directly on an analysis of this approach. For a more detailed exposition of the comparative strengths and weaknesses of Why-Because Analysis, the interested reader is directed to Johnson and Holloway (2003).

Systems Theory Accident Modelling and Process (STAMP)

STAMP is motivated by the observation that elements of systems theory might be applied to support the analysis of incidents and accidents. Mishaps occur when external disturbances are not adequately controlled. Similarly, adverse events can arise when the failure of process components goes undetected or when the actuators that might respond to such a failure are unsuccessful in their attempts to control any adverse consequences from the initial fault. Control failures can also arise from 'dysfunctional interactions' between system components. For example, this can arise if one subsystem embodies inappropriate assumptions about the performance characteristics of another process component. In this view, mishaps do not stem from events but from inappropriate or inadequate constraints on the interactions among the elements that form complex, safety-critical applications. Safety is viewed as a dynamic property of the system because the constraints that are applied and the degree to which a system satisfies those constraints will continually evolve over time.

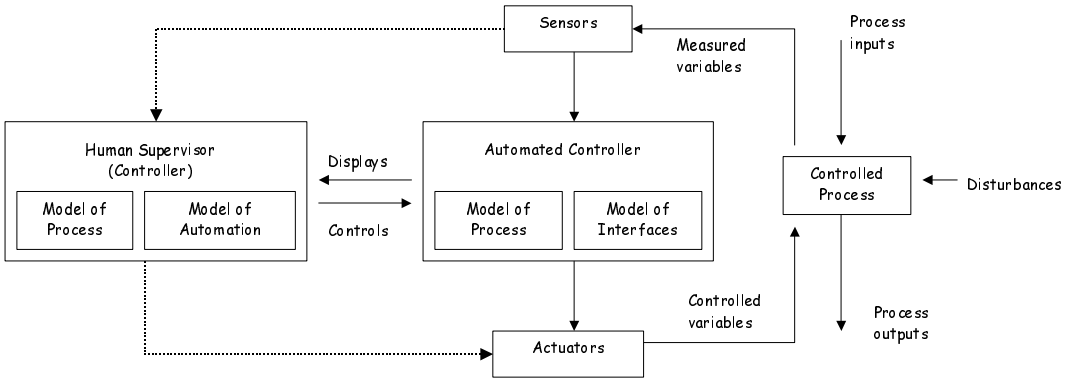

Figure 1: High-level Elements of a Control Model

Leveson has also used Figure 1 to extend the standard analysis presented in the previous paragraph. She argues that mishaps can also occur when operators have inappropriate internal models either of the process that is being controlled or of the automation that they use to interact with that process. For instance, if an automated system has already introduced a catalyst without the operator realising this then they may attempt to force the automated system to needlessly repeat this operation. Conversely, if the operator has an inadequate model of the controlled process then they might introduce a catalytic element at a time that might endanger future production. Similar comments can be made about the models that are embedded within a control system. This is significant because programmers, typically, must make assumptions about the processes that their software will help to control. Further problems can occur when control system designers assume that they will be able to obtain information from sensors within time constraints that cannot always be met. Such incorrect assumptions are characterised by an inadequate model of the interface between the control system and the controlled process, in Figure 1.

This form of analysis can be extended upwards from the operator, the control system and the production process to consider the relationships between project and company management, between management and regulatory agencies and between regulation and legislature. These different relationships must be captured in any analysis because they have a profound influence on both the development and operation of safety-critical systems. After having conducted this extended form of control analysis, the STAMP technique progresses by considering each of the control loops that are identified in the 'socio-technical system'. Potential mishaps stem from missing or inadequate constraints on processes or from the inadequate enforcement of a constraint that contributed to its violation. Table 1 illustrates the general classification scheme that guides this form of analysis. It provides a classification scheme that helps to identify potential causal factors in the control loops that exist at different levels of the management and operation hierarchy characterised using diagrams similar to that shown in Figure 1. Leveson (2002) points out that the factors identified in Table 1 can be applied at all levels, however, the interpretation will differ. For instance, a failure in a sensor to provide the operator with information when they need it can be classified as a time lag leading to inadequate feedback. Similarly, the same classification can be used to describe the failure of company management to provide adequate information about a potential hazard to senior company executives.

1. Inadequate Enforcements of Constraints (Control Actions)
1.1 Unidentified hazards
1.2 Inappropriate, ineffective or missing control actions for identified hazards
1.2.1 Design of control algorithm (process) does not enforce constraints
- Flaws in creation process
- Process changes without appropriate change in control algorithm (asynchronous evolution)
- Incorrect modification or adaptation.
1.2.2 Process models inconsistent, incomplete or incorrect (lack of linkup)
- Flaws in creation process
- Flaws in updating process (asynchronous evolution)
- Time lags and measurement inaccuracies not accounted for
1.2.3 Inadequate coordination among controllers and decision makers
2 Inadequate Execution of Control Action
2.1 Communication flaw
2.2 Inadequate actuator operation
2.3 Time lag
3. Inadequate or Missing Feedback
3.1 Not provided in system design
3.2 Communication flow
3.3 Time lag
3.4 Inadequate sensor operation (incorrect or no information provided)

Table 1: Control Flaws Leading to Hazards (Leveson, 2002)

This section has introduced the STAMP analysis technique that is intended to address the perceived weaknesses in more traditional 'chain of events' approaches to incident analysis. As mentioned, there are relatively few examples of this technique being applied to analyse complex, software-related mishaps. The team that helped to develop the technique are also responsible for the existing case studies (Leveson and Allen, 2002). We were, therefore, motivated to see whether we could replicate the results that have been described for this approach. We were also interested to see whether two independent analysts would achieve similar results if they applied this approach to the same case study. A number of methodological problems complicate our analysis. In particular, STAMP is not supported by a detailed methodology. There is no requirement that the control modelling be conducted before any classification of control flaws can begin. Leveson (2002) simply argues that the control structure surrounding the accident needs to be understood before any analysis is conducted. Pragmatic constraints, including those of project management, forced us to develop a methodology to support our application of the STAMP technique. We, therefore, distinguished between two stages. The first focuses on the development of control models, similar to that shown in Figure 1. The second focuses on the classification and analysis of constraint 'failures' between agents in the control models using the taxonomy provided in Table 1. Different results might be achieved if STAMP were used in a different fashion. For example, analysts might iterate between modelling and classification. This more sustained form of analysis might be feasible for larger investigation teams working on high-consequence mishaps.

The remainder of the paper analyses the mission interruption that affected the operation of the Solar and Heliospheric observatory (SOHO). The SOHO mission is a joint venture between the European Space Agency (ESA) and the US National Aeronautics and Space Administration (NASA). The spacecraft was intended to carry a payload of scientific instruments prepared by nine European and three American research teams. The entire mission involved more than 200 scientific investigators who work in the areas of helioseismology, solar atmosphere and corona research and investigations into solar wind phenomena. The spacecraft was built by a team from Matra Marconi Space and was launched by NASA in December 1995. ESA was responsible for spacecraft procurement and for the final integration and testing. NASA was responsible for the launch, ground system support and in-flight operations. Communications and tracking were performed using the Deep Space Network. Allied-Signal Technical Services Corporation conducted the day-to-day operations under a contract from NASA's Goddard Space Flight Center. More information on the development and operation of SOHO is available from http://soho.nascom.nasa.gov.

The mishap that forms the focus for this paper occurred in June 1998 when contact was lost with the spacecraft after approximately two years of successful operation. The incident was preceded by a planned calibration of three gyroscopes and by a momentum management manoeuvre. The gyroscope calibrations involve checking the gyro output values when the spacecraft is known to have no rotational motion around its roll axis. Any residual value indicates that the gyro's output has drifted over time. Flight operations can then calculate an offset that must be applied to the gyro readings to obtain a true value. The spacecraft's Attitude Control Unit computer largely controls momentum management. The management activity is performed every two months so that the momentum imparted by reaction wheels can be reduced to tolerable limits. The speed of these reaction wheels can be used during flight to maintain the spacecraft's attitude in the face of external disturbance torques.

The board identified that there were no failures on the spacecraft but that a number of ground based errors during the performance of these scheduled tasks combined to create the mishap. Following the calibration, Gyro A was despun and deactivated. This had been introduced as a modification to previous practice and was intended to extend its operational life. However, Gyro A played an important role in a safety-net procedure known as Emergency Sun Reacquisition (ESR). A further error in the software controlling the calibration process left Gyro B showing a high-gain value. This indicated that the roll rate of the spacecraft was twenty times greater than it actually was. This resulted in on-board fault detection software triggering Emergency Sun Reacquisition. This event became known as ESR5, the 5th since launch and occurred at 19:16 EDT on June 24th 1998. The ESR event led to an automatic reconfiguration of the Gyros. Gyro A replaced Gyro C for ESR thruster-based control while Gyro B continued to be used for fault detection. The error in Gyro B's gain was detected and corrected. However, the fact that Gyro A was still despun was not identified. Using values from Gyro A, the spacecraft began roll thrusters firing. In less than a minute, Gyro B triggered another Emergency Sun Reacquisition process (ESR6) at 22:35, June 24th 1998. The

spacecraft was still oriented towards the sun and had power. However, its attitude was increasingly unstable. Ground personnel identified that Gyro B disagreed with Gyro A and so decided to deactivate Gyro B, which had been working correctly after the initial high-gain had been detected. Ground operations commanded the spacecraft into Initial Sun Acquisition mode and thrusters firing commenced again. The intention was to counteract the roll that was indicated by the despun Gyro A. Gyro B and the associated fault-detection was now disabled. ESR control was now no longer stable and telemetry was lost at 12:42 on June 25th. This may have been due either to communications loss or to a loss of solar and battery power.

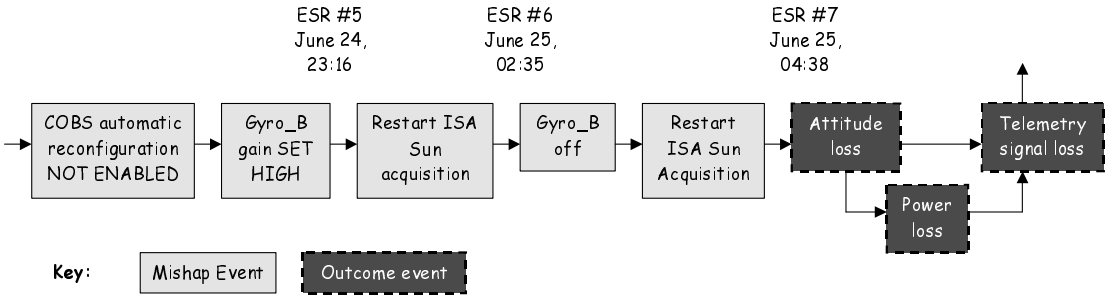

Figure 2: Failure Event Tree, Top Level View (NASA/ESA, 1998)

A joint investigation board was formed from NASA and ESA personnel. NPG 8621.1, mentioned in previous paragraphs, guided the investigation. In consequence, their analysis was supported by the use of Failure Event Trees. Figure 2 illustrates the high level view of the mishap that they developed using this technique. As can be seen, Failure Event Trees embody many of the ideas that are common to 'chain of events' models. A sequence of events leads to the mishap. These are denoted by the solid rectangles. Outcomes are denoted by rectangles drawn with dotted borders. Key timings, where they are known, are presented at the top of the diagram. The COBS acronym, in the initial mishap event, refers to Control On-board Software. This might have automatically reactivated all of the Gyroscopes during the ESRs.

Figure 3 provides a more complete overview. It illustrates how a series of managerial and organisational factors can be represented within Failure Event Trees. As can be seen, the investigation team traced the causes of some of the mishap events to improper code verification. The decision to despin Gyro A was introduced into the operating procedures without sufficient checking to ensure that this would not compromise the ESR mode. Similarly, the verification was insufficient to identify that Gyro B was set correctly following the calibration. Figure 3 identifies a number of technical issues that contributed to the verification failure. For instance, both ESA and NASA maintained their own simulations. Neither could be used to diagnose the error that might exhibit the symptoms that were being observed during the mishap. Figure 3 also identifies a range of management issues. These ranged from the lack of system knowledge possessed by some members of the Flight Operations Team through to the pressures that were imposed on Flight Operations by the Science teams. Flight Operations were forced to answer requests from the Science team during the mishap and this reduced the amount of time that they could devote to understanding the problems as they developed.

The extended Failure Event Tree also characterises a range of higher-level organisational issues. For example, the roles of ESA, NASA and Allied Signals staff were not clearly defined. Similarly, there were no higher levels of ESA or NASA management available when the mishap first occurred. This deprived the operations team of a valuable source of support. Had this support been available then the operations team might have been less concerned over interventions that could affect the scientific objectives of the project. Similarly, the presence of higher-level management might have encouraged operational staff to follow the procedures and analyse the incident more thoroughly before taking the decision to despin Gyro B. Finally, the investigation boards introduce an 'indirect factor' in Figure 3 to represent NASA pressure to reduce the costs associated with ground operations in order to support the extended life of the SOHO mission. This may have contributed to the high workloads of the operations team, which may, in turn, have created the circumstances for the mission interruption.

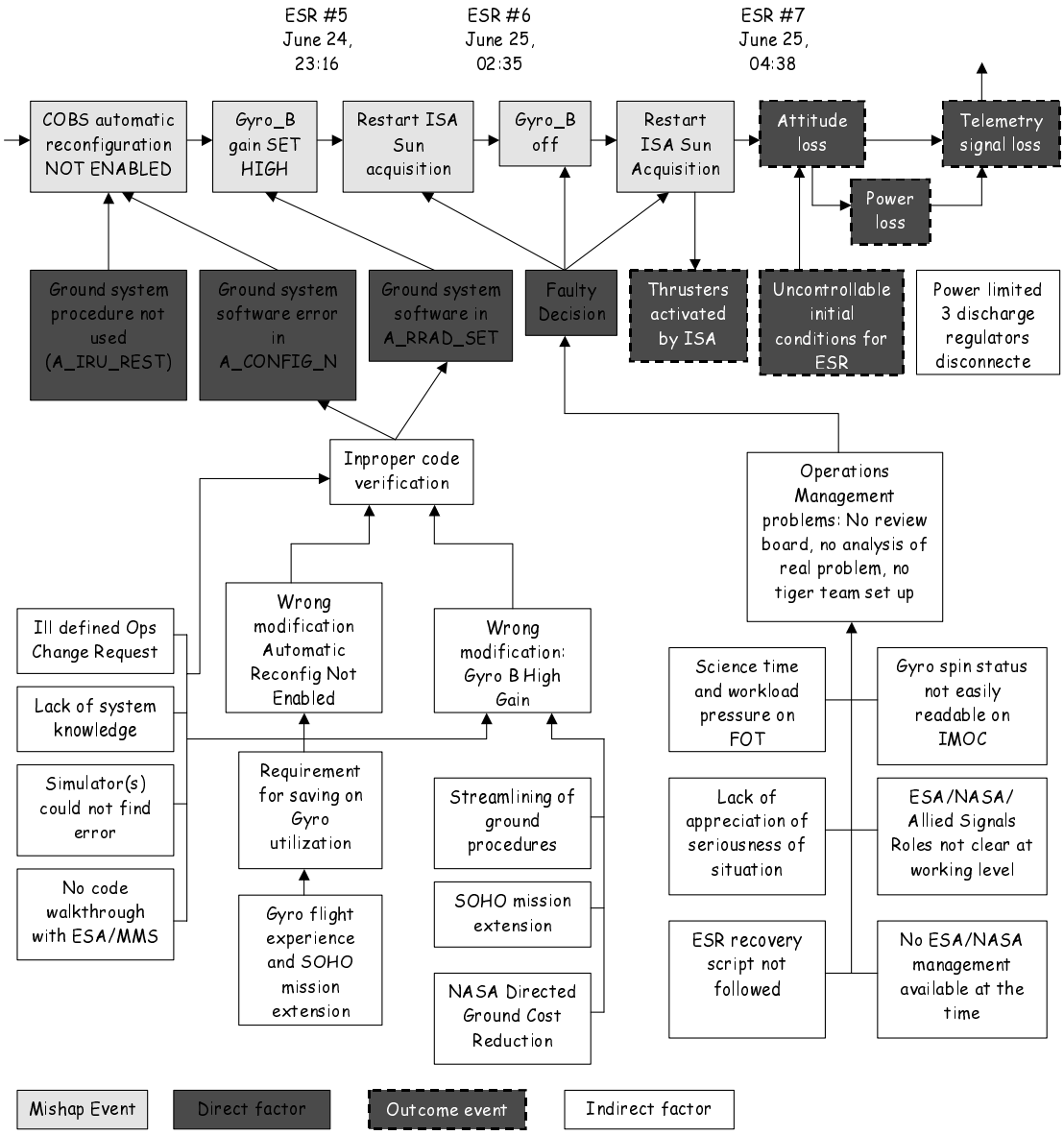

Figure 3: Failure Event Tree, Detailed View (NASA/ESA, 1998)

Leveson and Allen (2002) make a number of criticisms about 'chain of event' techniques, including the Failure Event Trees used by the Mishap Investigation Board. It can be difficult to determine the scope of any investigation. The selection of an initial event is often the result of an arbitrary decision. Investigators using these techniques will often disagree about the initial events that create the preconditions for a mishap to occur. Some 'chain of events' models, such as Failure Event Trees, often fail to distinguish between different types of events. Figure 3 contains missing process elements, such as the lack of a review board. Others nodes represent missing controls, including the lack of input from Gyro B prior to restarting Sun acquisition. The arrows between events introduce further confusion. They represent causal relationships. For example, a lack of systems knowledge during the reprogramming can be seen as one cause of the failure to perform adequate code verification. Elsewhere they represent the 'flow of events' without any causal information. For instance, the problems in reading the Gyro spin status are linked to the operational management problems in Figure 3. It could equally be argued that these problems were an effect rather than a cause of such managerial problems. Other techniques avoid these limitations.

Why-Because Analysis embodies formal ontologies within a rich type system. A limitation with this approach is that it can be complex and time consuming to maintain such explicit structures. The STAMP approach provides a midway point. It does not provide a formal ontology but does avoid the confusion between temporal and causal information, described above. We were, therefore, concerned to determine whether this approach would help to identify or emphasise causal factors that were different from those presented by the joint investigation team. We were also concerned to see whether two independent analyses would yield similar results using the STAMP control model approach to analyse a software related mishap. The following sections, therefore, document the application of this technique to the SOHO mission interruption.

STAMP Stage 1: The Control System Analysis

The first stage in our STAMP analysis identified the components of the control model that lies at the heart of this technique. As mentioned, we were concerned to determine whether individual differences might have an impact on the results of an investigation. The authors of this paper, therefore, used the joint investigation report as source material. We each constructed an initial model of the relationships between the systems, individuals and organisations that contributed to the interruption of services. The STAMP models provided in Leveson (2002) and Leveson and Allen (2002) were used as guidance during this stage of the investigation. Figures 4 and 5 illustrate the different approaches that were taken to this modelling activity.

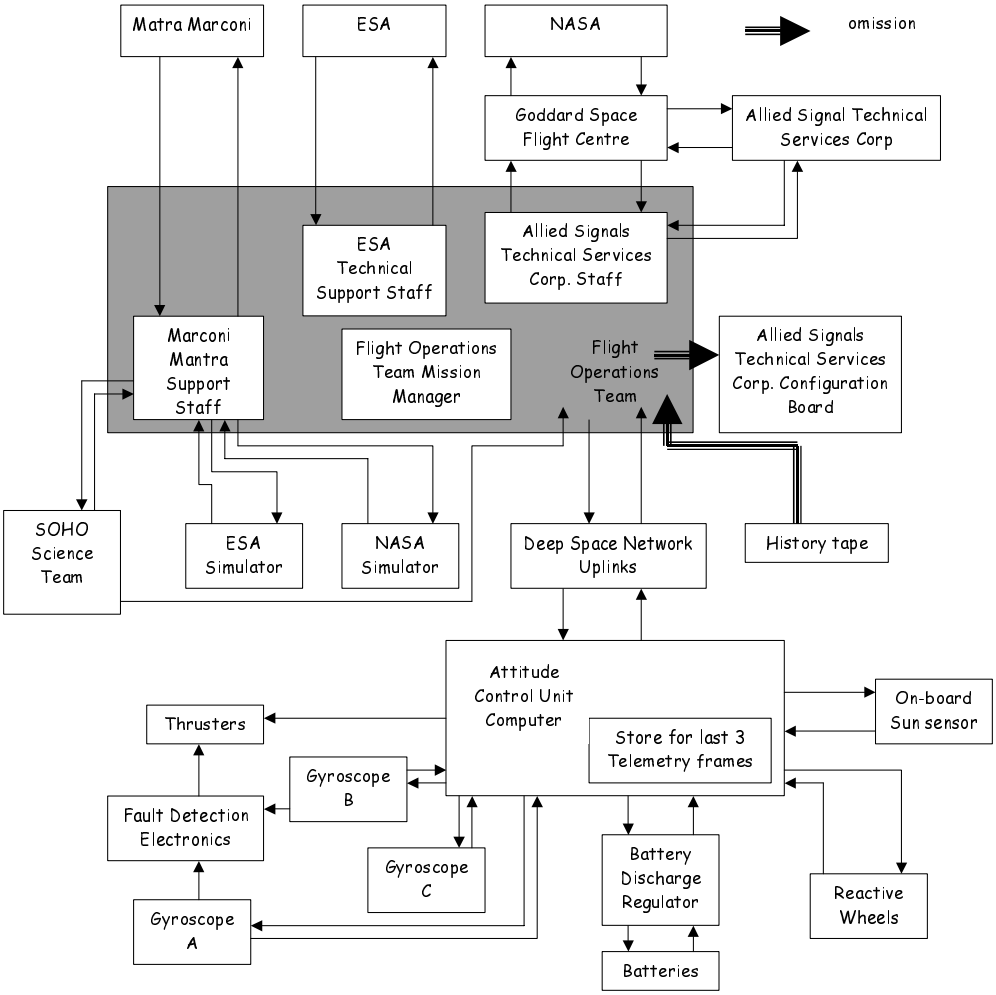

Figure 4: First STAMP Analysis of SOHO Mishap

Figure 4 illustrates a relatively detailed approach. It includes many of the relationships that are only implicitly captured in the Failure Event Trees of Figures 2 and 3. For instance, it distinguishes between the NASA and ESA simulators because the comparison of different results from these systems ate into the time that was available for mishap diagnosis. The STAMP technique also resulted in some components being looked at more closely than is apparent in the Failure Event Trees. For example, the Attitude Control Unit computer had sufficient storage to record the last three telemetry frames. The ground team could have used this information to help them identify the potential problem with Gyro A. Similarly, Figure 4 captures the relationship between the on-board batteries and the battery discharge regulators. Power may have been lost to the spacecraft in the aftermath of the mishap because the regulators were not operational. The control-based approach helps to make these issues explicit because it does not focus attention narrowly on the chain(s) of events leading up to the mishap itself. Hence, analysts are encouraged to look at the systems that were involved in the mitigation or exacerbation of the initial incident.

Figure 4 also includes a number of minor modifications to the control models that are sketched in the initial publications on STAMP (Leveson 2002, Leveson and Allen, 2002). A form of bold, double arrow is used to indicate control components that were not used. In particular, the history telemetry that was stored by ground-based systems was arguably not considered in enough detail before key decisions were made. Similarly, an Allied Signals Technical review board was not held to consider all of the software modifications mentioned in previous sections. We were initially concerned that the developers of the STAMP approach would reject the introduction of this additional annotation, which is already a feature of Why-Because Analysis (Johnson and Holloway, 2003). It can be argued that such control flaws should be considered during any subsequent classification, based on the taxonomy in Table 1. Figure 4 pre-empts this analysis, not for theoretical reasons but for pragmatics. These omissions were considered to be particularly salient during the initial analysis and hence were explicitly represented in the control model. This illustrates an important strength of the STAMP approach because the focus on mission controls encourages analysts to think about 'non-events', which are of course difficult to represent in 'chain of event' approaches. Our concerns about the introduction of the double arrow to indicate an unused control proved to be unwarranted. We provided Leveson with an early draft of this paper. She supported our use of the annotation and has since indicated that this syntactic device may be incorporated into future versions of the technique.

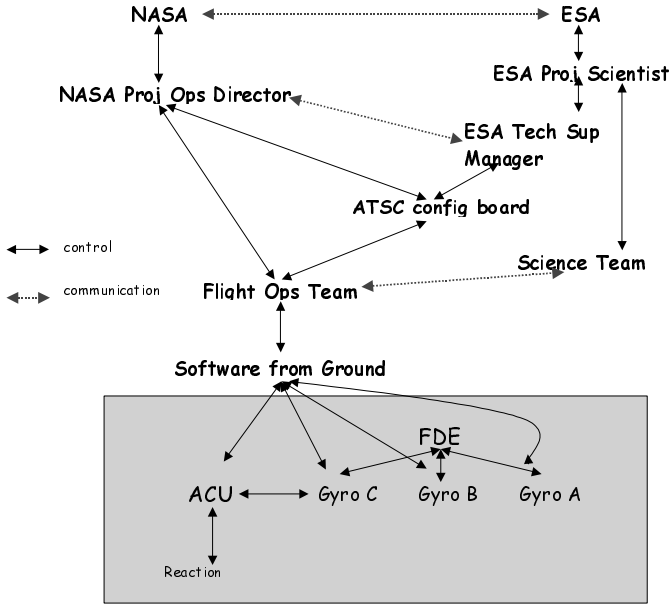

Figure 5: Second STAMP Analysis of SOHO Mishap

Comparison of our Results with the Previous Investigation

It is difficult to make precise claims about any 'added value' that was obtained from the control models beyond the insights provided by the initial NASA investigation. Our analysis was conducted after having read the first report. The control models are, therefore, strongly influenced by the previous investigation. This is a common problem in the assessment of causal analysis techniques. There is an understandable reluctance to trial new techniques in 'real' mishap investigations. Post hoc investigations cannot easily recreate the gradual elicitation of evidence and the iterative formation of hypotheses that characterise many enquiries (Johnson, 2003).

Having raised this caveat, it is possible to identify a number of differences between the control models and the official report. By focussing the analysis on a number of key actors in this mishap, Figure 5 clearly indicates those agents that the investigator considers to be central to the analysis. In particular, this control models recognises the significance of the configuration board. Figure 4 also independently recognises the importance of this board. However, the Failure Event Tree in Figure 3 includes this as past of a more general indirect factor; 'Operations management problems: no review board, no analysis of real problem, no tiger team set up'. This is surprising given the wider prominence of configuration management within the causes of the SOHO mission interruption:

> "… Though some of these modifications were made at the request of the SOHO Science Team, they were not necessarily driven by any specific requirement changes. The procedure modifications appear to have not been adequately controlled by ATSC configuration board, properly documented, nor reviewed and approved by ESA and/or NASA… It was only later realized that three of four battery discharge regulators were disconnected from the bus. In fact, analysis indicated that this condition had occurred several months prior and no one had recognized this change in the spacecraft configuration. This limited access to battery discharge current when the spacecraft needed it for control…" (NASA/ESA, 1998).

Figure 4 also illustrates differences in emphasis between the control models that were produced in our investigation and the arguments presented within the official NASA report into the SOHO mission interruption. It explicitly considers the role played by the history tapes and the store for the last three frames of telemetry data. At key moments during the mission interruption, these information resources could have been used to achieve a more accurate diagnosis of the state of the SOHO systems. However, neither of these features is mentioned in the Failure Event Tree of Figure 3. This arguably provides a further illustration of the need to consider non-events or the failure to use particular systems during the analysis of adverse events. This omission is all the more surprising given that the official investigation recognised the importance of these system components:

> "Real time data becomes limited because an autonomous data format change occurs whenever the spacecraft enters safe mode. However, the spacecraft status immediately preceding a safe mode trigger could have been determined by viewing the history tape that was generated prior to an anomaly. In addition, SOHO was designed to store, within its on-board computer for diagnostic purposes, the last three telemetry frames that preceded a safe mode entry. The operations script specifically states that the Gyro A is to be spinning upon entry into safe mode and instructs the operator to evaluate the three telemetry frames that had been stored prior to the anomaly before proceeding toward recovery. Neither the confirmation of Gyro A state nor the evaluation of the three previous telemetry frames was performed. These omissions resulted in a failure to notice that Gyro A was not spinning - a state which rendered the safe mode unstable. This ultimately resulted in a misinterpretation of Gyro B data, and eventually caused the loss of a SOHO safety net." (NASA/ESA, 1998).

This analysis illustrates how elements of the analysis in the official report cannot easily be traced back to the Failure Event Tree that is intended to drive the analysis. The STAMP approach is intended to ensure that such arguments are tied more closely to the underlying control models. This example illustrates further differences in emphasis between the STAMP control models and the initial report. Figure 4 represents the history tape and the telemetry frames as key components in the system architecture. In contrast, the previous citation comes from a section of the NASA report on 'Factors indirectly contributing

to failure; Part B. Procedure Implementation', paragraph 5 on 'Failure to follow the operations script; failure to evaluate primary and ancillary data'.

Analysis of the STAMP Method

Figure 5 presents the second control model that was produced during our analysis of the SOHO mission interruption. This diagram was developed by the second co-author of this paper working independently from his colleague. Both analysts had access to the same original, source material. There are, however, a number of clear differences between the two models. For instance, Figure 5 distinguishes between communication and control in a manner that is not apparent in Figure 4. There are further differences in the level of detail shown in the two models. Figure 5 provides a high-level overview of the key agents involved in the mission failure.

The differences between these two models raise a number of questions about the practical application of the STAMP approach. It might be argued that Figures 4 and 5 reflect the lack of guidance that is provided on the development of appropriate control models for mishap analysis. If this were the case then a set of heuristics could be devised that might help investigators derive more consistent models. In order for this to be successful, further research would need to identify the guidelines that are intended to encourage consistency. It is unclear precisely what heuristics might guide investigators towards control models that are more likely to yield 'significant' insights into the causes of an adverse event.

Alternatively, it might be argued that these different models reflect differences in the investigators' expertise in applying the STAMP approach. Both, however, had access to the same documentation on the technique and had attended the same introductory presentations. These diagrams reflect their initial attempts to apply the technique. It is entirely possible that subsequent experience in the application of STAMP might encourage greater convergence. It is difficult to make any firm conclusions on this point given that STAMP is a relatively new technique and is also undergoing further development.

In debriefing sessions, it transpired that each of the investigators had different intentions behind the models that they had developed. The simpler approach in Figure 5 was intended to introduce the salient features of the mission interruption. Figure 4 was intended to support a more detailed form of analysis in which a greater selection of 'actors' was explicitly represented. It can, therefore, be argued that both Figure 4 and Figure 5 might be simultaneously maintained during an investigation because they serve different purposes. The high-level model supports an exposition of the analysis to individuals and groups outside the immediate investigation team. A more detailed model may, however, also be necessary to support the fine-grained analysis of an adverse incident. The simultaneous development of two control models at different levels of granularity creates a number of problems. In particular, tool support is necessary to ensure consistency between these multiple representations. Figures 4 and 5 illustrate some of the problems that might arise should this form of support not be available. For example, Figure 4 refers to a Flight Operations Team Mission Manager while Figure 5 refers to NASA Project Operations Director. Further analysis was required to determine that different individuals performed these distinct roles.

STAMP Stage 2: Reliability Analysis of Constraint Classification

One of the aims of this research was to determine whether STAMP might help investigators to independently arrive at similar conclusions about the causes of a mishap. The initial results in Figures 4 and 5 reveal similarities between the objects, people and organisations that are represented in the control models. It remains to be seen whether this would be sustained if both versions were developed to the same level of detail. We were also concerned to determine whether the subsequent stages of the STAMP approach also encouraged consistency between different analysts. There are several ways of performing such comparisons. We might have continued the independent control model analysis into a subsequent classification of constraint 'flaws'. This was rejected because it was thought unlikely that two independent teams would pursue the simultaneous analysis of the same incident. Instead, we conferred to identify a common set of control related issues that arose during our independent analysis of the SOHO mishap. These can be broadly compared to thirteen of the causal factors that were identified by the original joint investigation team.

Each of the authors then independently assigned each of these factors to one of the items in the classification taxonomy of control flaws introduced in Table 1. The results were collated for later comparison. For instance, both analysts identified an *unintended hazard (category 1,1)* in the 'failure to perform risk analysis of a modified procedure set'. However, one observer classified 'dilution of observatory support' as an example of *inadequate actuator operation (category 2,2)* while the other classified it as *inadequate sensor operation (category 3,4)*. This illustrates further insights into the STAMP technique. Both interpretations of this control flaw are justified. The removal of observatory support reduced the ability of flight operations to both intervene and to monitor critical information. This is precisely what one might expect from the model of human supervisory control illustrated in Figure 1. Operators play a role in both monitoring and intervening in the control of process variables. However, neither of the analysts recognised this dual role until after they had compared their analysis. One reason for this is that we had decided to assign each causal factor to a single element of the constraint 'flaw' taxonomy. The complex, multifaceted nature of the causal factors that had been revealed and the high-level nature of the taxonomy initially justified this decision. Several of the causal factors could reasonably have been assigned to three or four of the elements in Table 1. Future applications of the STAMP might, therefore, benefit from additional support to guide investigators in their analysis of constraint 'flaws' if consistent classifications are to be sustained when each causal factor is assigned to several different elements in the taxonomy.

We took an alternative approach that was based narrowly on the pragmatics that guided our analysis. We lacked the time and resources to perform a cross comparison of multiple, independent classifications for each causal factor. In contrast, we decided that there was no definitive 'correct' assignment and that the independent analysis yielded better results when we met back together to discuss our findings. These discussions did not encourage us to change our initial classifications. Instead they persuaded us that there were good reasons why we each, separately identified different causes of the same control flaw. This finding has important pragmatic implications. If a causal factor can be assigned to more than one category then each analyst must consider whether every factor can potentially be assigned to every category. The associated overheads rapidly increase with the number of causal factors and elements in the classification. We believe that this is a necessary price to be paid for 'higher-risk' incidents that have severe potential consequences or a likelihood of future recurrence.

Analysis of the STAMP Method

Table 2 summarises the results of our analysis. We were surprised by these findings. At one level there might appear to be considerable disagreement. This is apparent when comparisons are made between the most detailed granularity in the classification. However, there is considerable agreement at the top-most level between inadequate enforcement of control actions, inadequate execution of control actions and inadequate or missing feedback. We did consider performing a statistical analysis of these results. However, we felt that the sample was too limited. Further studies are required to see whether such general levels of agreement are sustained for different analysts over a broader selection of mishaps. It is important to emphasise that both of the analysts had worked together on previous mishap analysis projects and both were working with identical sets of documentation. Set against these factors that might encourage agreement is the observation that one has a background in systems engineering whilst the other is a human factors specialist and computer scientist. To a limited extent, our study suggests that STAMP may help to overcome the biases that Lekberg (1997) has identified between analysts from different backgrounds. We did not, however, prioritise which of the control flaws were the most significant! Leveson in her response to an initial draft of this paper has argued that this is entirely appropriate and reflects the need to separate the understanding of why accidents occurred from the steps that are taken to prevent them. She argues that there is no scientific basis for such a prioritisation and that there are legitimate disagreements about what the goals of such a prioritisation should be. For instance, regulators may have a different perspective than line-management. It can, however, be argued that it should be possible to trace the constraint failures that motivate any subsequent allocation of finite resources. Without this form of traceability, there is a danger that recommendations may be made without any firm foundation in the STAMP analysis. One way of resolving this apparent conflict would be to reject any automatic means of deriving recommendations from the application of STAMP but to support some form of syntax for documenting which part of the control analysis led to the identification of a particular set of recommendations. It would then be possible for

different investigators to recommend different measures from the same analysis whilst at the same time providing the documentary evidence that is needed to justify any expenditure on process improvements.

	Analyst 1	Analyst 2
1 Inadequate Enforcement of Constraints		
1.1 Unidentified hazards	Factor 2: "Failure to perform risk analysis of a modified procedure… Factor 4: "Failure to properly respect autonomous Safe Mode triggers	Factor 2: Failure to perform risk analysis of a modified procedure set. Factor 7: Failure to recognise risk posed by operations team overload
1.2 Inappropriate, ineffective or missing control actions for identified hazards		Factor 6: Failure to Question Telemetry discrepancies
1.2.1 Design of control algorithm does not enforce constraints	Factor 1: Flight operations team modified flight-demonstrated ground operations procedures as a part of the ISTP Ground System re-engineering… Factor 10: Over reliance of flight operations team on ESA and MMS representatives…	
1.2.2 Process model inconsistent, incomplete or inaccurate	Factor 6: Failure to Question Telemetry discrepancies Factor 9: Emphasis on science return at expense of spacecraft safety Factor 13: Failure to validate the planned sequence of events in advance.	Factor 1: Flight operations team modified flight-demonstrated ground operations procedures as a part of the ISTP Ground System re-engineering… Factor 9: Emphasis on science return at expense of spacecraft safety
1.2.3 Inadequate coordination among controllers and decision makers	Factor 7: Failure to recognise risk posed by operations team overload Factor 8: Failure to recognise shortcomings in implementation of ESA/NASA agreements…	Factor 10: Over reliance of flight operations team on ESA and MMS representatives…
2 Inadequate Execution of Control Actions		
2.1 Communication flaw		Factor 5: Failure to follow the operations script; failure to evaluate primary and ancillary data…
2.2 Inadequate actuator operation	Factor 5: Failure to follow the operations script; failure to evaluate primary and ancillary data… Factor 11: Dillution of observatory engineering support…	Factor 12: Failure to resolve a critical deficiency report in a timely manner Factor 4: Failure to properly respect autonomous Safe Mode triggers
2.3 Time lag	Factor 12: Failure to resolve a critical deficiency report in a timely manner	
3. Inadequate or Missing Feedback		
3.1 Not provided in system design		Factor 13: Failure to validate the planned sequence of events in advance.
3.2 Communication flaw	Factor 3: Failure to communicate change	Factor 3: Failure to communicate change
3.3 Time lag		
3.4 Inadequate sensor operation		Factor 8: Failure to recognise shortcomings in implementation of ESA/NASA agreements… Factor 11: Dillution of observatory engineering support…

Table 2: Comparison of Analyst's Constraint Classification

Comparison of our Results with the Previous Investigation

Table 3 illustrates the classification of causal factors that is adopted in the ESA/NASA report on the SOHO mission interruption. As can be seen, there are several noticeable differences between this allocation and those embodied within Table 2. Leveson's taxonomy is entirely general hence individual causal factors are allocated to headings such as 'unidentified hazards' rather than domain specific topics such as 'ground procedures' or 'ground systems'. One consequence of this is that the STAMP analysis groups causal factors under common categories that were not considered together within the official report. For instance, Factors 5 and 12 are both related to inadequate execution of control actions in the STAMP analysis but are related to procedure implementation and ground systems in the official report.

Categories in Mishap report	**Factors contributing to Mishap**
A. Ground procedures	1. Flight operations team modified flight-demonstrated ground operations procedures as part of the ISTP Ground System reengineering…
	2. Failure to perform risk analysis of a modified procedure…
	3. Failure to communicate change
B. Procedure Implementation	4.Failure to properly respect autonomous Safe Mode triggers.
	5. Failure to follow the operations script; failure to evaluate primary and ancillary data.
	6. Failure to question telemetry discrepancies.
C. Management structure and process	7. Failure to recognise risk posed by operations team overload.
	8. Failure to recognise shortcomings in implementation of ESA/NASA agreements.
	9. Emphasis on science returns at expense of spacecraft safety.
	10. Over-reliance on flight operations team on ESA and MMS representatives.
	11. Dilution of observatory engineering support.
D. Ground Systems	12. Failure to resolve a critical deficiency report in a timely manner.
	13. Failure to validate the planned sequence of actions in advance.

Table 3. Allocation of Factors in the NASA/ESA (1998) report

The differences between the NASA/ESA and STAMP stem from different objectives. Leveson's methodology is intended to provide a general structure for incident analysis. The constraint taxonomy used in Table 2, therefore, reflects gross distinctions between classes of control failure. In contrast, the official report uses the same classification to group both causal factors and the subsequent recommendations:

> **RECOMMENDATIONS A. GROUND PROCEDURES**
> 1. An ESA and NASA review of the process for SOHO operational procedure change should be implemented forthwith. The review should critically assess the process from beginning to end. The review should include matters such as who can initiate a change, who agrees it should be made, how is the modification process monitored, how is it validated, how is it introduced into operations, how it is signed-off, how are the users of the procedure informed of the change, and how are users trained on the new version...

Similar paragraphs consider the recommendations that are directed at procedure implementation, management structure and process, and ground systems. It is difficult to be certain about the consequences that these different classifications would have upon the outcome of a mishap investigation. It might be argued that the generic taxonomy in STAMP requires an additional stage of analysis so that causal factors are grouped in terms of the particular groups that must implement them. Ground systems requirements might be distinguished from management structures in the manner exploited by the official report. Conversely, it can be argued that the specific classification of Table 3 must be generalised to address the wider problems that are illustrated by this specific mishap. For example, the STAMP analysis identified that Factors 1, 6, 9 and 13 all related to situations in which process models were inconsistent, incomplete or inaccurate.

Arguably the most important aspect of any comparison between STAMP with existing NASA/ESA investigation techniques would focus on the different recommendations generated from these approaches.

Unfortunately, we were not able to complete this analysis. As noted previously, STAMP deliberately excludes any methods for deriving interventions from the products of a causal analysis. This forms a strong contrast with other techniques, including Why-Because Analysis, that use various forms of counterfactual reasoning to help identify potential recommendations (Johnson, 2003). In private correspondence, Leveson has justified this decision by pointing out the difficulties of finding evidence to support counterfactual arguments of the form 'the mishap would have been avoided if X had not occurred'. Her arguments are supported by the work of cognitive psychologists, most notably Byrne and Handley (1997). They have identified a range of common biases that affect this form of reasoning. It remains to be seen whether these criticisms affect more formal incident investigation techniques. In practical terms, however, both investigators felt the need for some form of methodological guidance in using the outcomes from the STAMP causal analysis to generate recommendations.

Results

The previous section identified a number of methodological concerns that were raised by this research. In addition to the caveats that we have already discussed, a number of further issues should be recognised. We did not perform an independent analysis of constraint 'flaws'. We identified and agreed upon a subset of the concerns identified by the original joint investigation board. We did this because our expertise is in the software engineering, systems development and the application of mishap analysis techniques. We are not domain experts. Further concerns stem from the observation that STAMP is a relatively new technique. There is not a large body of literature or case study material. Hence, we may have made mistakes in our application of the approach. Measures were taken to avoid this by close reference to published papers (Leveson, 2002, Leveson and Allen, 2002). We also sent a preprint of the paper to the developers of the STAMP approach to determine whether we had made any fundamental mistakes. The main technical criticism of our application of STAMP was that we had not focussed enough on the constraints that form an important component of the approach. Leveson (2002) argues:

> "A more appropriate way to understand the role of software in accidents is to use systems theory. In systems theory terminology, safety is an *emergent property* that arises when the system components interact within an environment. Emergent properties are controlled or enforced by a set of constraints on the system states that result from component interactions; for example, valve A must always be open whenever valve B is closed. Accidents result when safety constraints are violated. When software acts as a controller in such systems, it embodies or enforces the constraints by controlling components and their interactions (e.g., opening and closing the valves). Software, then, can contribute to an accident by not enforcing the appropriate constraints on behavior or by commanding behavior that violates the constraints".

This is an important criticism. We could, for example, have investigated more thoroughly the reasons why software modifications did not result in a configuration board meeting. The violation of this constraint is noted in the control model of Figure 4 but was inadequately addressed in our classification of Table 2. It is important to recognise, however, that our analysis and Leveson's subsequent critique provide mutual benefits. Not only do they encourage us to reconsider our analysis of the SOHO mission interruption but we have also motivated the developers of STAMP to reiterate the importance of constraint identification in their guidance material.

The most obvious benefit from using STAMP compared to 'chain of events' models is that it actively encourages investigators to consider the systemic causes of adverse events. The development of the models, illustrated in Figures 4 and 5, encourages investigators to look at the flow of information and control. As the technique embodies elements of traditional control engineering, these diagrams provide a convenient means of analysing the behaviour of complex technological systems. This is often difficult in event based techniques. The extension of these approaches to consider organisational relationships and human behaviour is beneficial. Our analysis illustrated that the combined representation of technical systems and managerial structures was convenient and appropriate for our case study.

The declarative nature of the STAMP modelling proved to be a useful corrective to the more procedural aspects of 'chain of events' models. By this we mean that the representation of organisational and managerial structures encouraged our analysis to look well beyond the catalytic events that triggered the

mishap. In contrast, we were encouraged to investigate the relationship between NASA, ESA and Allied Signal personnel. We were also encouraged to consider the interaction between key members of the Flight Operations Team, including management and the Marconi Matra engineers' interaction with the scientific teams.

STAMP also offers numerous benefits for the process of mishap analysis. It is relatively simple and easy to follow. We learned how to use the approach in a relatively short period of time. Both analysts believed that the technique could credibly be used by domain experts and by incident investigators with minimal training. However, there is a need for more guidance on the interpretation of the constraint 'flaw' classification that forms the second stage of the analysis. We were, however, impressed by the level of agreement that was obtained from our high-level classification of control flaws in the SOHO mishap report.

This research also identified a number of potential problems in the application of STAMP to our case study. The first is that both analysts needed to construct a 'chain of events' prior to constructing the control models. Our analysis was driven by a series of informal notes and diagrams that helped us to organise and make sense of the source material. The complexity of this and similar adverse events makes us concerned that entirely doing away with a rough sketch of the flow of events over time would lead to important omissions and ambiguities in our interpretation of what happened rather than why it occurred. This does not contradict the guidance offered by the proponents of STAMP who argue that 'chain of events' models can still be useful in determining 'what' occurred but provide few insights into 'why' those events happened. Equally, however, our analysis would have benefited from additional guidance on the translation from more convention 'timelines' to the control models in Figures 4 and 5.

A second concern focussed on the changes that occurred within the system that we studied. If we had conducted a full and detailed analysis of this mishap we would have been forced to draw and redraw figures 4 and 5 several times. For example, the Gyros were reconfigured so that their control and information flows changed. Similarly, we would have had to redraw the control models to reflect managerial changes. In this case study we were faced with the particular problem of whether to draw the management and organisational structure that was nominally supposed to exist or the one that actually evolved as a result of management changes during the operation of the mission:

> "At a lower level, the SOHO ESA/NASA Mission Management Plan further states that following in-orbit commissioning, mission management authority would be transferred from the ESA Project Manager to the ESA Project Scientist, resident at GSFC. The ESA Project Scientist would be supported by an ESA technical support manager also resident at GSFC, who would have access to additional ESA and MMS technical support. In practice after the spacecrafts first year in orbit, the support team was typically comprised of the ESA technical support manager, one MMS engineer and occasionally additional engineers from Europe for important manoeuvres…." (page 10, ESA/NASA 1998).

We can address the dynamism that characterises many adverse events by constructing multiple control models that each represents a snapshot of the structure of the system. These would be indexed over time and, hence, it can be argued would introduce elements of the 'chain of events' back into the control model. This goes against the underlying philosophy of the STAMP approach and would significantly increase the overheads associated with a relatively simple and cost effective approach.

Further concerns stem from the way in which STAMP models organisational structures. We wondered whether the same techniques might be applied to apparently successful organisations and yet yield similar results. For example, many safety-critical organisations explicitly adopt a range of flexible communications channels between different groups. This inevitably leads to some inconsistency between the information that is held about their mutual activities. However, this need not endanger the safety of an application process providing that communication takes place when necessary. The control flaw classification, in contrast, encourages analysts to identify such nominal behaviours as potential problems. Similarly, it seems highly unlikely that any two operators will ever maintain a mutually consistent model of any complex application process. No individual's mental model of an application will ever be complete except at the highest level of abstraction. These caveats stem from the way in which STAMP seeks to

extend the same constraint 'flaw' criteria form automated control systems to human operators and dynamic organisations. It is possible to do this but many theoretical and pragmatic objections remain to be addressed.

Finally, our experience in applying STAMP raised a number of concerns about the integration between the two stages of the analysis. There was a tendency to identify the control flaws without reference to the more detailed models, especially that shown in Figure 4. One means of avoiding this problem would be to provide more guidance on the methodology needed to extract the constraint 'flaws' from the models prior to classification. For example, the classification might be used to identify a series of heuristic questions that would guide the identification of particular flaws. This approach has been adopted within the EUROCONTROL guidance on human error analysis (Isaac, Engelen and Polman, 2002).

Conclusions and Further Work

Mishap investigations provide important information about adverse events and are intended to help avoid any recurrence of previous failures. However, the complexity of many safety-critical software-related mishaps poses new challenges for accident analysis. Similarly, the recognition that many failures have complex, systemic causes has helped to widen the scope of incident investigations. A new generation of mishap analysis techniques has been proposed to help investigators address these problems. For instance, Leveson (2002) has recently developed the STAMP approach to address some of the weaknesses associated with previous 'chain of event'. There are relatively few examples of the STAMP technique. This paper, therefore, presented the results that were obtained when two investigators used this approach to analyse the software-related causes of a mission interruption to the joint ESA and NASA Solar and Heliocentric Observatory (SOHO).

Our analysis identified significant benefits from using STAMP. In particular, it provides a useful corrective to 'chain of event' models. It achieves its objective of encouraging a more systemic approach to mishap investigation. The approach is relatively easy to use and is arguably cost effective in terms of the insights that are obtained for the effort involved in performing the analysis. We have also identified a number of problems that complicated our application of STAMP. The most important of these are that it is difficult to apply the 'constraint' flaw taxonomy to complex organisations and human operators. A control modelling approach is also extremely difficult to use when complex, technological systems allow for dynamic reconfiguration and when organisational structures slowly evolve away from a documented management structure.

Future studies should replicate our work with other analysts looking at a broader spectrum of incidents. We are also particularly concerned to identify ways in which STAMP might be better integrated with other stages of the mishap investigation process. At present, there is no guidance on how to gather the evidence that supports a control models or the constraints that it embodies. This is particularly important considering the previous caveats about explicitly representing dynamic organisations. How do we find out what was actually happening within a management structure rather than what was supposed to be happening?

A further and final concern is that some link must be made between the products of a STAMP classification and the recommendations that are intended to avoid future mishaps. At present, it provides powerful support for identifying the shortcomings of complex systems. It does not enable analysts to identify those higher priority control flaws that must form the focus for future investment and regulatory intervention. As we have noted, Leveson rejects this role for STAMP. She insists that there is no systematic means of deriving such prioritisations. We have, therefore, argued that future work might develop techniques for documenting the relationship between constraint 'flaws' and any subsequent recommendations. The intention is that analysts and regulators might then be able to identify the arguments that justify particular interventions. The developers of the STAMP approach view it as a means of integrating mishap investigation with existing forms of risk analysis. We envisage that this form of methodological support might help to link these techniques so that the analysis of previous failures can inform future development.

Acknowledgements
This research has been funded by a NASA contract NAS1-97046, Task 212. UK Engineering and Physical Sciences Council grant (EPSRC GR/M98302) has provided additional support. Thanks are due to Peter Bishop (Adelard) and Nancy Leveson (MIT) who provided valuable feedback on an early draft of this paper. The authors would also like to thank the referees who proposed useful corrections.

References

R.M.J. Byrne and S.J. Handley (1997), Reasoning strategies for suppositional deductions, Cognition, 62,1-49.

A. Isaac, P. Engelen and M. Polman, (2002), Human Error in European Air Traffic Management: From Theory to Practice. In C.W. Johnson (ed.), Investigation and Reporting of Incidents and Accidents 2002. Department of Computing Science, University of Glasgow, Scotland.

IEC 61508, (2000) Functional safety of electrical/electronic/programmable electronic safety-related systems, International Electrotechnical Commission. See http://www.iec.ch/61508 for further details.

C.W. Johnson (2003 in press), A Handbook for the Reporting of Incidents and Accidents, Springer Verlag, London, UK.

C.W. Johnson and C.M. Holloway, (2003), A Survey of Logic Formalisms to Support Mishap Analysis, Reliability Engineering and System Safety, 79, accepted and to appear.

P. Ladkin and K. Loer (1998), Why-Because Analysis: Formal Reasoning About Incidents, Bielefeld, Germany, Document RVS-Bk-98-01, Technischen Fakultat der Universitat Bielefeld, Germany.

A.K. Lekberg, (1997), Different Approaches to Incident Investigation: How the Analyst Makes a Difference. In S. Smith and B. Lewis (eds.) Proceedings of the 15th International Systems Safety Conference, 178-193, Systems Safety Society, Unionville, VA, United States of America.

N. Leveson, (2002), A Systems Model of Accidents. In J.H. Wiggins and S. Thomason (eds) Proceedings of the 20th International System Safety Conference, 476-486, International Systems Safety Society, Unionville, USA.

N. Leveson and P. Allen, (2002), The Analysis of a Friendly Fire Accident Using a Systems Model of Accidents. In J.H. Wiggins and S. Thomason (eds) Proceedings of the 20th International System Safety Conference, International Systems Safety Society, Unionville, USA.

NASA/ESA, (1998), SOHO Mission Interruption Joint NASA/ESA Investigation Board Final Report. Available from http://umbra.nascom.nasa.gov/soho/SOHO_final_report.html

NASA (2001), NASA Procedures and Guidelines for Mishap Reporting, Investigating and Record-keeping,Safety and Risk Management Division, NASA Headquarters, NASA PG 8621.1, Washington DC, USA, http://www.hq.nasa.gov/office/codeq/doctree/safeheal.htm.

J. Reason (1997), Managing the Risks of Organizational Accidents, Ashgate Publishing, Aldershot, UK.