ASHGATE PUBLISHING LTD

Advances in Aviation Psychology

Chapter 2

Applying Systems Thinking to Aviation Psychology

Nancy G. Leveson

Massachusetts Institute of Technology, USA

Hazard analysis is at the heart of system safety. It can be described succinctly as "investigating an accident before it happens." A hazard is selected, such as two aircraft violating minimum separation standards or an aircraft losing sufficient lift to maintain altitude, and then the scenarios that can lead to that hazardous state are identified. Hazards are informally defined here as precursor states to accidents that the designer never wants the system to get into purposely. The resulting scenarios or potential paths to the hazard are then used to compute the probability of the hazardous state occurring or to design to either eliminate the scenarios or to control or mitigate them. Alternatively, after an accident, hazard analysis techniques can generate the potential scenarios to assist accident investigators in determining the most likely cause.

Most of the current widely used hazard analysis methods were created 50 or more years ago when the systems being built were simpler and were composed primarily of electro-mechanical components. Human operators mostly followed pre-defined procedures consisting of discrete and cognitively simple tasks such as reading a gauge or opening a valve. Failure rates and failure modes could be determined through historical usage or through extensive testing and simulation. Humans were either omitted from these calculations or were assumed to "fail" in the same way that electro-mechanical components did, that is, randomly and with an identifiable probability. Safety engineers and human factors experts existed in separate worlds: the safety engineers concentrated on the hazardous scenarios involving the physical engineered components of the system and human factors experts focused on the human operator such as training and the design of the physical interface between the human and the engineered system.

As software was introduced to increase functionality and desired system properties (such as efficiency and fuel savings), the role of the operator changed from one of direct controller to supervisor of the automation that actually flew the plane. The increasing complexity led to new types of human error (Sarter & Woods, 2008) and stretched the limits of comprehensibility for both the designers and the operators of these systems. We are now designing systems in which operator error is inevitable, but still blame most accidents on the pilots or operators. Something then is either done about the operator involved, such as fire them or retrain them, or engineers do something about operators in general, such as marginalizing them further by automating

more control functions or rigidifying their work by creating more rules and procedures, many of which cannot be followed if the system is to operate efficiently (Dekker, 2006).

At the same time, the hazard analysis methods were not updated to take into account the new types of accident scenarios that were occurring and to treat the operator as an integral part of the larger system. As a result, hazard analyses often miss possible scenarios, especially those involving software or humans. To make progress, we need the psychology, human factors, and engineering communities to come together to create more powerful hazard analysis methods—and therefore ways to improve the system design—that are appropriate for the systems being built and operated today. This chapter describes a potential approach to doing that. It starts from an extended model of accident causality called STAMP (System–Theoretic Accident Model and Processes) that better describes the role humans and software play in accidents today (Leveson, 2012).

In the next section, STAMP and an associated new hazard analysis method called System–Theoretic Process Analysis (STPA) are described along with the resulting implications for more sophisticated handling of humans in engineering analysis and design. Proposed changes to ATC (NextGen) are used as an example. Then open questions are described in which the aviation psychology community could provide important contributions.

How Are Accidents Caused?

Traditional safety engineering techniques are based on a very old model of accident causation that assumes accidents are caused by directly related chains of failure events: failure A leads to failure B which causes failure C, which leads to the loss. For example, the pitot tubes freeze, which causes the computer autopilot to stop operating (or to operate incorrectly), followed by a stall warning that is incorrectly handled by the pilots, which leads to the plane descending into the Atlantic. This chain of events is an example of an accident scenario that might be generated by a hazard analysis. The underlying model of causality implies that the way to prevent accidents is to prevent these individual failure events, for example, train pilots better in how to react to a stall warning and improve the pitot tube design.

The chain-of-events model served well for simpler systems, but our more complex, software-intensive systems are changing the nature of causality in accidents. Software does not fail randomly and, in fact, one could argue that it does not fail at all. Software is an example of pure design without any physical realization. How can an abstraction fail? It certainly can do the wrong thing at the wrong time, but almost always accidents related to software are caused by incorrect requirements, that is, the software engineers did not understand what the software was supposed to do under all conditions, such as when false readings are provided by the pitot tubes.

In the same way, human contributions to accidents are also changing, with the rise in importance of system design factors, such as mode confusion, that cannot be explained totally by factors within the human but instead result from interactions between human psychology and system design. Many accidents today are not caused by individual component failure but by unsafe and unintended interactions among the system components, including the operators.

The STAMP model of accident causality was created to deal with the new factors in accidents and to consider more than individual or multiple component failure in causal analysis (Leveson, 2012). Accidents are treated not as a chain of component failure events but as the result of inadequate enforcement of constraints on the behavior of the system components. In this case, the system includes the entire socio-technical system. Figure 2.1 shows an example of a typical hierarchical safety control structure in aviation. Each component in the structure plays a role in accident prevention and, therefore, in accident causation. The control structure on the left ensures that safety is built into the system (for example, aircraft) and the control structure on the right ensures that the systems are operated safely. There are usually interactions among them. Each of the components in Figure 2.1 has a set of responsibilities or safety constraints that must be enforced by that component to prevent a hazard.

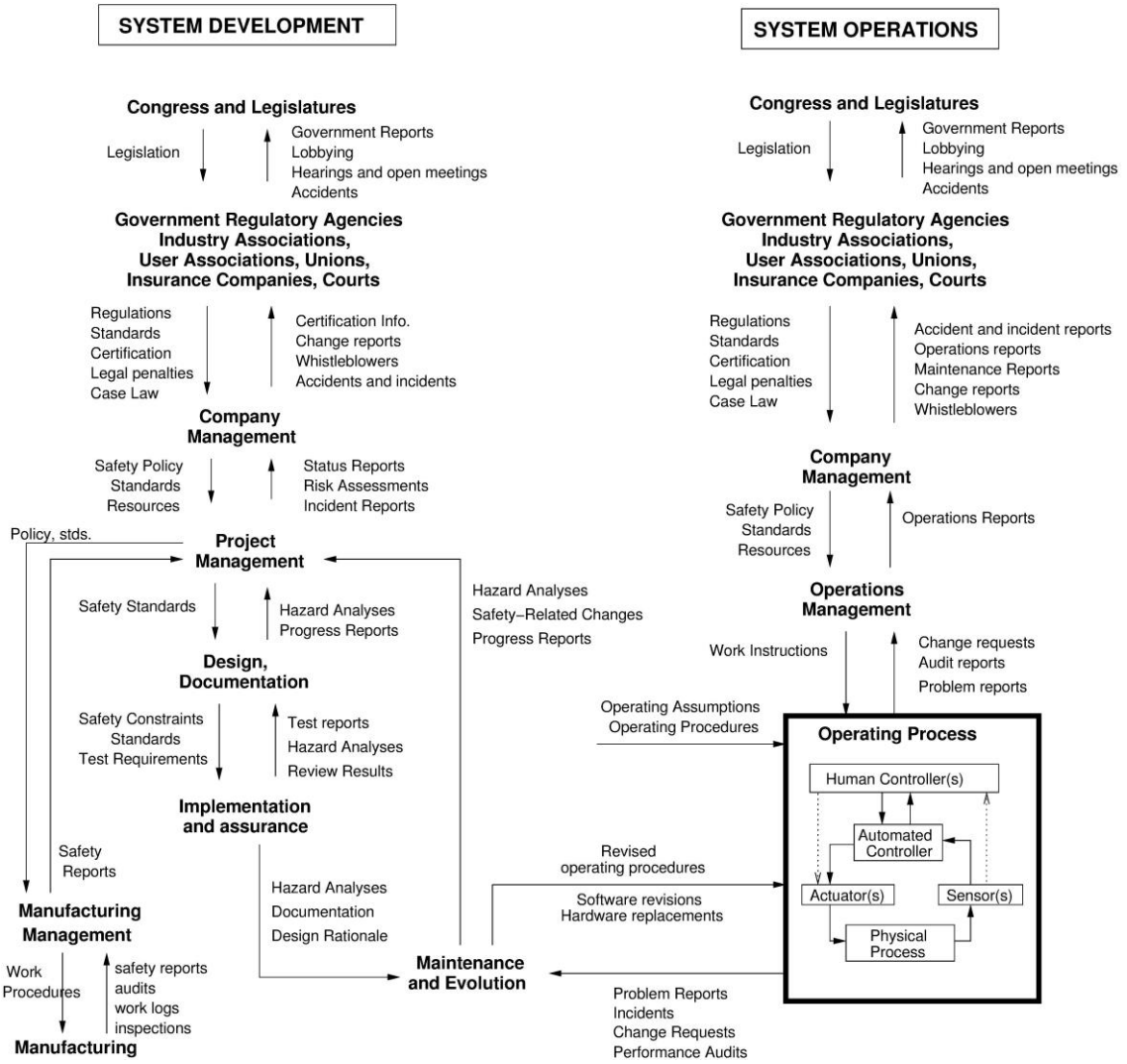

Figure 2.1 An example of a hierarchical safety control structure

Figure 2.2 shows the hierarchical control structure (omitting the upper levels for simplicity) involved in a new ATC procedure called In-Trail Procedure (ITP) that allows aircraft to pass each other over the Atlantic airspace even though minimum separation requirements may be violated temporarily during the maneuver. Information about the location of both aircraft is provided through Global Positioning System (GPS) and ADS-B and the ITP equipment onboard the aircraft determines whether passing will be safe at this point. If the ITP criteria for safe passing are met, the pilots can request a clearance to execute the maneuver. A hazard analysis of this system would attempt to generate the scenarios in which ITP could lead to an accident. That information can then be used by engineers and human factors experts to try to prevent accidents either through system design changes or operational procedures.

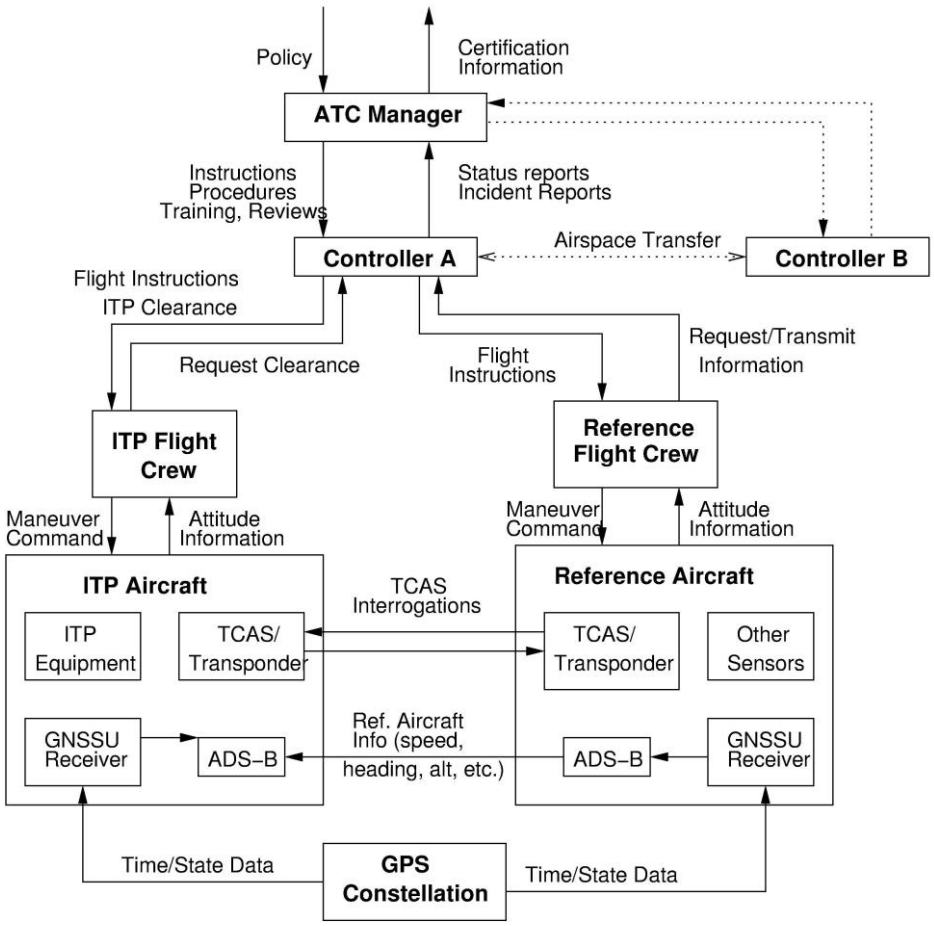

Figure 2.2 The safety control structure for ITP

An important component of STAMP is the concept of a *process model* (see Figure 2.3). The safety control structure is made up of feedback control loops where the controller issues commands or control actions to the controlled process, for example, the pilot sends a command to the flight computer to ascend. In order to operate effectively, every controller must have a model of what it thinks is the state of the subsystem it is controlling. The actions or commands that the controller issues will be based at least partly on that model of the state of the system. If this model is incorrect, that is, inconsistent with the real state of the system, then the controller may do the "wrong" thing in the sense it is the right thing with respect to the information the controller has but wrong with respect to the true state of the system. If the pilots or the ATC controller has an incorrect understanding of whether the criteria for safe execution of the ITP are met, for example, they may do the wrong thing even though they have not themselves "failed" but simply were misled about the state of the system.

The process model is kept up to date by feedback and other inputs. In humans, the process model is usually considered to be part of the *mental model*. Note that the feedback channels are crucial, both in terms of their design and operation.

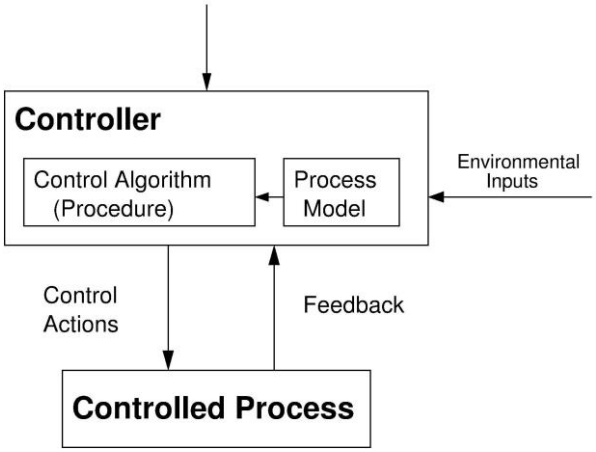

Figure 2.3 Every controller contains a model of the state of the controlled process

While this model works well for software and is certainly a better model for how humans work than that of random failure, it can be improved with respect to accounting for human factors in accidents. Some ideas for achieving this goal are presented later. But first, the implications of the present model are considered and the results of using it in hazard analysis compared with traditional hazard analysis methods.

There are four types of unsafe control actions that can lead to an accident.

1. A command required for safety (to avoid a hazard) is not given. For example, two aircraft are on a collision course and neither Traffic Collision Avoidance System (TCAS) nor an ATC Controller issues an advisory to change course.
2. Unsafe commands are given that cause a hazard. An example is an ATC Controller issuing advisories that put two aircraft on a collision course.
3. Potentially correct and safe commands are given, but at the wrong time (too early, too late, or in the wrong sequence). For example, TCAS provides a resolution advisory for the pilot to pull up too late to avoid a collision.
4. A required control command is stopped too soon or continued too long. For example, the pilot ascends as directed by a TCAS resolution advisory but does not level off at the required altitude.

Although classic control theory and control commands are emphasized here, the model is more general in terms of accounting for other types of controls on behavior than just a physical or human controller in a feedback loop. For example, component failures and unsafe interactions may be controlled through design using standard engineering techniques such as redundancy, interlocks, or fail-safe design. System behavior may also be controlled through manufacturing processes and procedures, maintenance processes, and operations. A third and important type of control over behavior comes through social controls, which may be governmental or regulatory

but may also be cultural values, insurance, the legal system, or even individual self-interest. The goal of design for safety is to create a set of socio-technical safety controls that are effective in enforcing the behavior required for safety while at the same time allowing as much freedom as possible in how the non-safety goals of the system are achieved.

Identifying Hazardous Scenarios

STPA is a new hazard analysis method based on the STAMP accident causation model. It works as a top-down system engineering process that starts with system hazards and then identifies behavioral constraints that must be imposed on the system components in order to ensure safety. It also assists safety analysts and system designers in identifying the set of scenarios that can lead to an accident. In practice, STPA has been found to identify a larger set of scenarios than found by traditional hazard analysis techniques, such as fault trees, event trees, and failure modes and effects analysis, particularly with respect to those scenarios involving software or human behavior (for example, Balgos, 2012; Fleming, Spencer, Thomas, Leveson & Wilkinson, 2013; Ishimatsu et al., 2014; Pereira, Lee & Howard, 2006).

To understand how STPA works, consider the ITP (In-Trail Procedure) example (RTCA, 2008). The STPA process first identifies the types of unsafe control actions that can lead to particular hazards and then uses that information and the control structure to identify the causes or scenarios that could lead to the unsafe control action. In the previous section, four general types of unsafe control action were identified. These are listed across the top of Table 2.1. The flight crew has two types of control actions they can provide (column 1): an action to execute the ITP and an action to abort it if they believe that is necessary. Within the table, the types of hazardous control actions are listed, for example, executing the ITP when the ATC Controller has not approved it or executing it when the criteria for safe passing are not satisfied. The actual process (along with automated support) to create the table are beyond the scope of this chapter but the reader should be able to see easily how this could be accomplished. A complete ITP analysis can be found in Fleming et.al. (2013).

Table 2.1 Potentially unsafe control actions by the flight crew

Controller: Flight Crew	Not Providing Causes Hazard	Providing Causes Hazard	Wrong Timing/Order Causes Hazard	Stopped Too Soon/Applied Too Long
Execute ITP		ITP executed when not approved. ITP executed when criteria are not satisfied. ITP executed with incorrect climb rate, final altitude, etc.	ITP executed too soon before approval. ITP executed too late after reassessment.	ITP aircraft levels off above requested FL. ITP aircraft levels off below requested FL.
Abnormal Termination of ITP	Flight crew continues with maneuver in dangerous situation.	Flight crew aborts unnecessarily. Flight crew does not follow regional contingency procedures while aborting.		

Once the unsafe control actions have been identified, their potential causes are identified using the generic types of failures or errors that could occur in the control loop as shown in Figure 2.4. The information about the potential causes can then be used for system design to eliminate or reduce them, create operational procedures, design training, and so on. For example, consider the reasons for why the flight crew might execute the ITP maneuver when it has not been approved or when the criteria are not satisfied. There are a lot of such reasons, but many are related to the flight crew's mental model, that is, they think that the approval has been given (when it has not) or they think the criteria are satisfied when they are not. Some scenarios involve the crew getting incorrect information, different sources give conflicting information, misperceptions about what information they have received, and so on. These scenarios (reasons) are used to design protection against the unsafe behavior by the flight crew and to create detailed requirements for the design of the system.

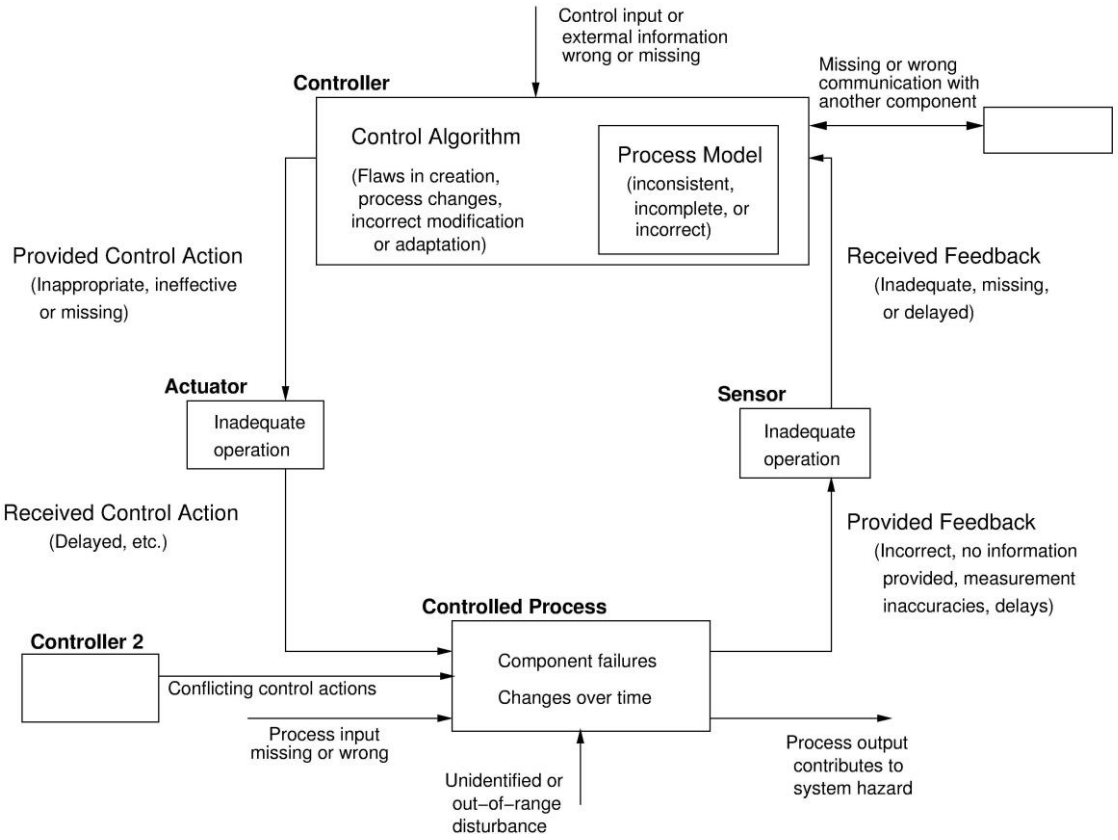

Figure 2.4 Generic types of problems in a general control loop that could lead to unsafe control

An important question, of course, is whether STPA is better than the traditional hazard analysis methods that are being used for NextGen. The official hazard analysis for ITP uses a combination of fault tree and event trees (RTCA, 2008). Probabilities are assigned to human error through a subjective process that involved workshops with controllers and pilots and eliciting how often they thought they would make certain types of mistakes.

As an example, one of the faults depicts the scenario for executing the ITP even though the ITP criteria are not satisfied. The fault tree analysis starts with an assigned probabilistic safety objective of 1.63e-3 per ITP operation at the top of the tree. Three causes are identified for the unsafe behavior which is approving an ITP maneuver when the distance criterion is not satisfied: (1) the flight crew does not understand what the ITP minimum distance is; (2) ATC does not receive the ITP distance but approves the maneuver anyway; or (3) there are communication errors (partial corruption of the message during transport). Probabilities are assigned to these three causes and combined to get a probability (1.010e-4) for the top event, which is within the safety objective.

The goal in the official risk assessment is to determine whether the maneuver will be within the assigned safety objective and not to improve the design. The fault tree analysis gives no guidance on how to prevent the human errors but instead assumes they happen arbitrarily or randomly. The fault tree also assumes independent behavior, but the interaction and behavior of the flight crew and ATC may be coupled, with the parties exerting influence on each other or being influenced by higher-level system constraints. Finally, the analysis asserts that communication errors are due to corruption of data during transport (essentially a hardware or software error), but there are many other reasons for potential errors in communication.

The STPA results include the basic communication errors identified in the fault tree, but STPA also identifies additional reasons for communication errors as well as guidance for understanding human error within the context of the system. Communication errors may result from confusion about multiple sources of information (for either the flight crew or ATC), from confusion about heritage or newly implemented communication protocols, or from simple transcription or speaking errors. There is no way to quantify or verify the probabilities of any of these sources of error for many reasons, particularly because the errors are dependent on context and the operator environments are highly dynamic. Instead of assuming that humans will rarely "fail," the STPA analysis assumes they will make mistakes and specifies safety and design requirements accordingly.

Possible Extensions to System–Theoretic Process Analysis (STPA) for Human Factors

While STPA as defined above is proving in a lot of comparative studies to be better than traditional hazard analysis techniques, it needs to be improved. The first step would be to provide a less naïve model of the human controller. While humans do not fail like mechanical components, they also do not operate with fixed algorithms (procedures) like computers as assumed above. Figure 2.5 shows a more realistic model of the role of humans in STAMP.

There are three levels of control shown in Figure 2.5. The bottom two, that is, the controlled process and an automated controller, are the same as shown previously. The top level is a first attempt at a more sophisticated model of the behavior of a human controller. Rather than having a fixed control algorithm (or procedure) that is always strictly followed, humans generate control actions using a model of the controller process, a model of the automated controller, a model of the context in which the control is taking place as well as written or trained procedures.

Leveson (2012) has identified some basic design principles using this model to reduce human controller errors, for example, ways to support the controller in creating and maintaining an accurate mental model of the controlled process and of the automation. Known problems,

such as mode confusion are included. While these design principles are not unknown in aviation psychology, they are restated in a way that engineers can apply them directly to their designs. These principles could and should be expanded.

Another important improvement would be to extend the STPA process to include more fundamental human factors concepts. The resulting analysis could have important potential implications for providing engineers with the information necessary to design systems that greatly reduce the types of human error contributing to accidents.

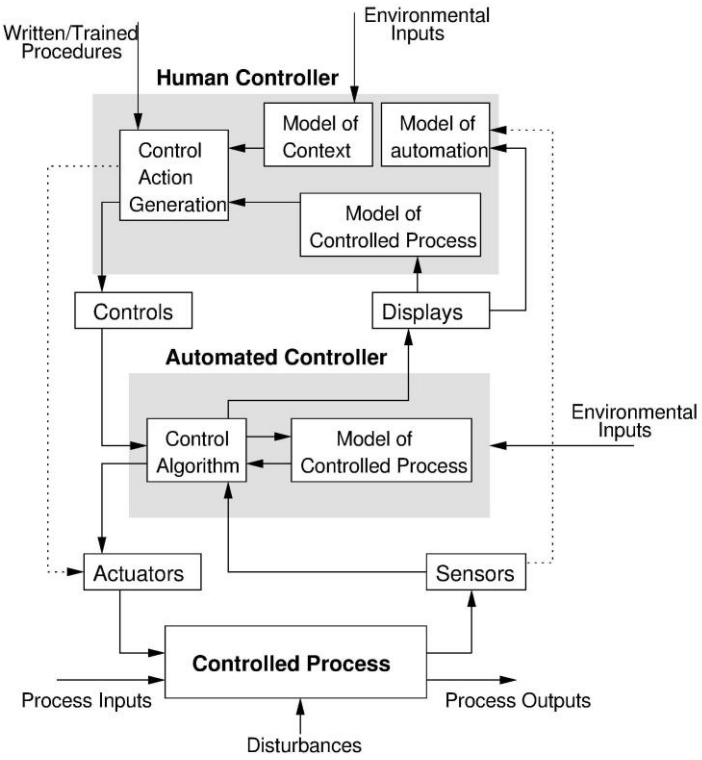

Figure 2.5 An extension to include a more realistic model of human behavior

Conclusions

Engineering needs to get beyond greatly oversimplifying the role of humans in complex systems but the aviation psychology community will need to help them. Hazard analysis and system design techniques that were created 50 years ago are no longer useful enough. This chapter has described a new, expanded model of accident causation, STAMP, based on systems thinking that could be the start for engineers and human factors experts to work together to create much safer systems.

Acknowledgements

This research has been partially funded by NASA Contract NNL 10AA13C. The author also learned much from conversations with John Thomas about human factors and STAMP.

References

Balgos, V. (2012), A systems theoretic Application to design for the safety of medical diagnostic devices, MIT Master's Thesis, February.

Dekker, S. (2006), *The field guide to understanding human error*, London: Ashgate.

Fleming, C.H., Spencer, M., Thomas, J., Leveson, N., & Chris Wilkinson (2013), Safety assurance in NextGen and complex transportation systems, *Safety Science, 55*(June), 173–187.

Ishimatsu, T., Leveson, N.G., Thomas, J.P., Fleming, C.H., Katahira, M., Miyamoto, Y., Ujiie, R., Nakao, H., & Hoshino, N. (2014), Hazard analysis of complex spacecraft using STPA, *AIAA Journal of Spacecraft and Rockets,* in press.

Leveson, N. (2012), *Engineering a safer world*, Cambridge, MA: MIT Press.

Pereira, S.J., Lee, G., & Howard, J. (2006), A system-theoretic hazard analysis methodology for a non-advocate safety assessment of the ballistic missile defense system, *Proceedings of the 2006 AIAA Missile Sciences Conference*, Monterey, CA, November.

Sarter, N., & Woods, D.D. (1995), How in the world did I ever get in that mode? Mode error and awareness in supervisory control, *Human Factors 37*(1), 5–19.

RTCA (2008), Safety, performance and interoperability requirements document for the in-trail procedure in the oceanic airspace (ATSA-ITP) Application, DO-312, Washington DC, June 19.