Modeling the 2005 Hatlestad Slide

Ying Qian

Shanghai University, School of Management

Leire Labaka Zubieta

University of Navarra, Industrial Management Department

Peter Lango

University of Bergen, Institute for administration and organization

Jose J. Gonzalez

Agder University, Centre for Integrated Emergency Management

ABSTRACT

It has long been recognized that the management of emergencies requires that response organizations act flexibly, becoming an "emergent organization" to better manage the fact that disasters do not follow scripts. Nevertheless, recent research shows that crisis response organizations prefer to follow patterns adequate for normal situations. Arguably, the resistance to become an emergent organization could be related to poor understanding of how to move from disorganization to self-organization. We present a system dynamics model describing the transition from disorganization to self-organization in the management of the fatal Hatlestad landslide in Norway. We suggest that the causal structure of the system dynamics model describing the Hatlestad case should be considered a candidate for an emergent "middle-range theory" describing the management of disorganization in emergencies. We propose specific data collection to test the candidate theory.

Keywords

Emergency management, management of disorganization, emergent organizations, theory building, system dynamics

INTRODUCTION

Disasters can be difficult to predict and often come by surprise. Emergency organizations can hardly prepare for all types of disasters. Dynes and Quarantelly (1976) pointed out from a compilation of 36 studies published 1957-1973 that organizations trying to hold on to traditional plans and procedures run into problems when an emergency occurs and do a worse job at responding than those who move to greater flexibility or resiliency. Despite this early and well-known insight, a recent study based on questionnaire responses from an international sample of emergency managers shows that "response organizations prefer to follow the general protocols of communication set-up for normal situations also in the event of a crisis" (Palttala, Boano et al. 2012, p. 5).

Crisis management is extremely complex, even becoming increasingly more complex (Boin and Hart 2008). Several theory-based approaches haven used and new approaches are emerging. Gilpin and Murphy (2008, p. 4) argue that the contingency, uncertainty, and happenstance; the unexpected confluence of unrelated events; and the destabilizing influence of rapidly changing circumstances are central factors in complex crises. Hence, an approach is needed that "maintains a vision of the changeable and complex nature of crises, and it looks for ways to operate within that very real-world environment of confusion, unforeseen events, and missing information."

Accordingly, complexity theory, "the study of interaction processes within complex systems, including social systems such as organizations" (Gilpin and Murphy 2008, p. 6) ought to play a central role to advance our understanding of crisis management. Gilpin and Murphy are not the only proponents of an approach based on complexity theory. Other recent research implies that crisis operations need to be "multi-organizational, trans-jurisdictional, polycentric response networks." They demand lateral collaboration and coordination, not top-down command and control (Boin and Hart 2008, p. 147).

Simpson and Hancock (2009) point out that most past work on emergency response has addressed well-structured problems of emergency services. Such studies assumed implicitly that an "Established" organization (Dynes and Quarantelli 1976) responds based on detailed planning. However, disasters tend to disrupt plans and impair the very same organization in charge of the emergency response. Simpson and Hancock analyze future research opportunities and conclude that “a single thread unites the particular areas of research opportunities discussed earlier in the paper: each is a problem of working in a disorganized environment” (Simpson and Hancock 2009, p. S136). They go as far as characterizing management of disorganization as an important operational research growth area for the next 50 years (Simpson and Hancock 2009, p. S126).

The case for a new approach based on complexity theory is strong. However, complexity theory is a vision involving many and intricate aspects related to interaction, holism, emergence and chaos (Byrne 1998). Related concepts are self-organization, complex adaptive systems, and co-evolution. Complexity theory attempts a generic, overall explanation of a whole range of phenomena – it is a *general theory* (in the sense of Schwaninger and Grösser, Schwaninger and Grösser 2008, p. 450).

How to apply the theory of complexity to the organizational behavior in emergencies has been little studied. In particular, computable models of crisis and emergency management using complexity theory are virtually absent. One needs a specific theory that link observables leading to models and simulations that can be compared to the behavior of real cases. In other words, a *middle-range theory* must be found. By a middle-range theory one understands “theories that lie between the minor but necessary working hypotheses that evolve in abundance during day-to-day research and the all-inclusive systematic efforts to develop a unified theory that will explain all the observed uniformities of social behavior, social organization and social change” (Merton 1968, p. 39).

System dynamics modeling is colloquially characterized by the system dynamics community as ‘theory building’. Schwaninger and Grösser (2008) studied whether system dynamics enables the construction of high-quality theories. They provided strong evidence for that using a set of criteria for high-quality theories on a revelatory case (in the sense of Yin 2009, p. 45ff). By analogy, the process of system dynamics modeling, as proposed in major text books (Richardson and Pugh 1981; Sterman 2000; Maani and Cavana 2007) and as practiced in high quality research work satisfies the denomination of theory building.

A recent article in the System Dynamics Review (Tu, Wang et al. 2009) applies complexity theory, in the sense of describing the emergence of order in dynamic nonlinear systems, to develop a system dynamics model of a the handling of a minor crisis – the Palau incident. The crisis is described and analyzed in the book “Cognition in the Wild” (Hutchins 1995). Hutchins data collection was performed on the vessel ”Palau”, on a four-day journey. On the last entry to the port the ship lost its power, and all electrical systems shut down. The navigation crew guided Palau into port despite the malfunctioning equipment. The crew had to operate on the verge of chaos, responding continuously to navigational cues obtained by calculating the ships position manually based on landmarks.

The Palau incident was a minor crisis, but it does share important characteristics with major crises and emergencies. Tu et al. argue that the handling of the Palau incident followed a common set of related events, disequilibrium–experimenting–emergence process as described by MacIntosh and MacLean (1999). This disequilibrium-experimenting-emergence process rhymes quite well with the empirically founded principle that successful handling of emergencies requires that the organization transitions to an Emergent organization (Dynes and Quarantelli 1976; Turoff, Gonzalez et al. 2012).

Tu et al.’s system dynamics model can be seen as a middle-range theory for the Palau incident. The authors show that the model renders the reference behavior of the emergency management as derived from the analysis provided by Hutchins (1995) and satisfies the canonical verification and validation principles of best-practice system dynamics modeling. Thus, Tu et al.’s work satisfies the criteria proposed by Schwaninger and Grösser (2008). In fact, Tu et al.’s work satisfies also this requirement: “For a theory what is required is a model along with a plausible account of why the model produces the behavior that it does” (Lane 2008).

Because of the relevance of Tu et al.‘s framework a replication study was conducted. Replication is extremely important to science (Jasny, Chin et al. 2011), but replication studies in social and management sciences were rare (Sterman 2000, p. 855ff) and probably are still rare. Our replication study, with kind help from Ya-tsai Tseng, the corresponding author in Tu, Wang and Tseng (2009), confirmed their main findings related to the emergence of order in dynamic nonlinear systems. However, the system dynamics model used for the results described by Tu et al. is very complex. Further, there were several errors in the paper, including a mismatch between the actual simulation model developed by the authors using ithink and the model diagrams in the paper, which were drawn with Vensim. The replication study could show that the system dynamics model could be significantly reduced in size and simplified, while still being able to render the reference behavior from the Palau case. Further we extended Tu et al.’s work and developed a preliminary system dynamics model of the Hatlestad Slide (Gonzalez, Bøe et al. 2013).

The work by Gonzalez et al. (2013) was preliminary on two accounts. First, the reference behavior modes are quite speculative since they were derived by generous extrapolations from the description of the Hatlestad slide in a master thesis (Lango 2010) and a book chapter (Lango 2011). Both sources, while rich in details, did not target a potential system dynamics simulation, so that the data therein that could be used to compose reference behavior for crucial variables is rather incomplete. Second, the study extends a (much simplified version) of Tu et al.'s paper (2009), which is not optimal, owing to the large differences between the Palau case (a minor incident) and the Hatlestad slide (a major tragedy).

In this work we develop a system dynamics model from scratch based on reference behavior modes obtained by interviewing first responders to the Hatlestad slide. The structure of the paper is as follows. First, the Hatlestad Slide is described. Second, we present the reference behavior obtained from interviews with first responders. Third, the structure of the system dynamics model is discussed. Fourth, a simulation and feedback analysis is conducted. Fifth, the model's simulation behavior is shown to agree quite well with the reference behavior modes and a feedback analysis is conducted. Sixth, we show that the model passes a series of extreme tests. Seventh, a sensitivity analysis is conducted. In the last section we discuss the obtained results and suggest that the system dynamics model could be considered a rudimentary middle range theory for emergency management.

THE HATLESTAD SLIDE

Hatlestad Terrasse is a neighborhood in the city of Bergen, Norway, consisting of housing units in a rural hilly setting. In the early morning of 14 September 2005 the cliff above the neighborhood broke apart, and a slide of clay, mud and rock hit a row of four houses while their residents were asleep. The slide went through the ground floors without causing the houses to collapse. People on higher floors were not harmed, and they could evacuate without need of assistance through windows in the first floors. But the residents in the ground floors were severely afflicted: ten people were buried and three died as a result, while seven more persons were wounded. A total of 225 people in the Hatlestad surroundings were evacuated (Wikipedia 2007).

The disaster investigation pointed to extreme weather conditions in combination with the steepness of the hill behind Hatlestad Terrasse as a major cause of the landslide (Multiconsult AS 2005). For several weeks before the disaster Western Norway experienced record-breaking precipitation. One weather station, Bergen-Florida, recorded the highest ever precipitation (156.5 mm/day during the fatal day of 14 September 2005) since the establishment of the weather station in 1875 (Meteorological Institute Oslo 2005).

The Hatlestad slide was the first of a series of slides in Norway. Shortly afterward a new slide occurred in Helebakken, Bergen. A new major slide occurred in Ålesund 2008 and in Namsos 2009. The total impact of the slides can be considered an agenda-setting crisis (Boin and Hart 2008, p. 19; Lango 2010, p. 91). The Hatlestad tragedy and the other slides in Western Norway were extensively covered in the media and had a strong impact on public opinion, so that the policies for housing construction on hills became stricter. Also, preparedness toward extreme weather changed. After the major slides, evacuation of citizens at risk is routinely considered in conditions of extreme weather. Other preventative measures were implemented, so for example the city of Bergen, which mapped out housing potentially at risk in case of extreme weather (Wikipedia 2007). Furthermore, firefighters were given new equipment and got training for this kind of situations.

REFERENCE MODES FOR THE HATLESTAD SLIDE EMERGENCY

In order to obtain the reference modes of key variables, several interviews were carried out with first responders that took part in the Hatlestad slide. Although these interviewees got a first-hand experience, it is important to note that these interviews were carried out 8 years after the tragedy and therefore, they had some difficulties in remembering all the details. Thus, at the beginning of the interview, the timeline of the Hatlestad slide was presented to help remember all the milestones that occurred.

Initially, the invitation was sent to nine first responders from different fields (police, health, civil defense, municipality and firefighters) to participate in the interviews. Nevertheless, only 4 experts agree to participate in the process (see Table 1). The reasons for not taking part in the process were different: some experts argued that they were no more working in the same position and therefore they preferred not to take part in the process. Others argued that they did not remember much about the tragedy and they did not feel able to answer the questions. The interviews were conducted in the autumn of 2013. All of the interviews took place at the interviewees' workplace.

Organization	**Role**
Hordaland Police District, Bergen Center	Emergency management, steering and coordinating
Hordaland Police District, Bergen South	Emergency management, steering and coordinating
The Civil Defense, Hordaland District	Supporting function
Municipality (Bergen), Section for Safety and Preparedness	Take care of people affected by the landslide both under and in the aftermath of the landslide

Table 1: Organizations of experts that took part in the interview

The interview had two main objectives: (1) to confirm that the already made hypotheses are correct and (2) to obtain reference modes to validate that the model represents the reality. In the first part, short questions were asked to the interviewees regarding some experienced situations to confirm some of the already set hypotheses. These questions were related to the availability of information during the emergency, amount of data during the emergency, understanding problems among different teams involved in the emergency, the need for improvisation, the causes of errors committed etc. The experts were asked to evaluate from 0 to 5 (being 0 no experience and 5 high experience) to what extent they experienced the proposed situations.

Experts stated that the following situations are the ones that they more experienced:

- *Misunderstandings with other teams involved in the emergency:* different stakeholders often use different vocabulary or they have different procedures or priorities when responding to a crisis. This can lead to misunderstandings and problems when making decisions about what has to be done.
- *Increase of workload:* when non-unpredicted events occur, these may generate high-stress situations. Beforehand established procedures may not be suitable and as a result, the workload of responders increases significantly above normal situations.
- *Need to improvise response actions:* related to the previous situation, when already defined procedures are not sufficient or suitable to respond, new response actions need to be improvised to properly response.
- *High stress situation:* every crisis creates stress situation and the non-unpredicted ones create even higher stress-situations. High-stress situations can lead to commit errors in performing tasks and making decisions.

In the second part of the interview, the experts were asked to draw the behavior over time graphs of key variables, which were identified from previous work (Tu, Wang et al. 2009; Gonzalez, Bø e et al. 2013). Bearing in mind that the meaning of some variables could be confusing, the survey provided a concrete definition for each variable. Furthermore, during the interview, the interviewer explained each variable informally using some examples to ensure that the meaning the expert was using corresponded to the one provided by us. Below, the analyzed key variables and the given description to each of them are explained:

- *Expected/Actual workload:* when a crisis occurs the workload of responders increases above the normal level. These variables measure the quantity of task that needs to be carried out during the emergency and the number of tasks actually performed in each moment, respectively.
- *Cognitive load:* when we are doing something very familiar, we use a small part of the mental capacity. On the other hand, when we are performing something very unfamiliar, or performing several activities at the same time, we might need to use more mental capacity.
- *Improvisations:* already defined procedures and protocols are often not suitable to cope with unpredictable situations. In these cases, improvisations have to be done, such as defining new initiatives or improvising emergent solutions to respond to the crisis.
- *Mismatch of understanding:* this variable represents the difficulties that different teams (stakeholders) in response activities may experience in understanding new decisions and actions that have been improvised locally by other members due to the extraordinary situation.
- *Quality of the decisions:* when facing an extraordinary situation, the decisions often have to be taken under stress situations and without much information. After the emergency, when looking back to the taken decisions we might see that they were not the most appropriate or efficient ones. Thus, this variable refers to how good or proper the decisions taken during the emergency situation are.

The aim of this exercise was to gather information about how these key variables evolve during the Hatlestad tragedy. This research was not interested in the concrete values but in the shape of the curves, that is, if the variables increase or decrease, how much and for how long these variations remain (using relative values).

The graphs obtained vary from one expert to another in some cases. The biggest difference lies in the graphs provided by the fourth expert since his role was quite different from the rest of the experts. He had a supporting role that means that his role was to take care of people affected by the landslide both under the management and in the aftermath of the landslide. So while the other actors were almost finished with their tasks the active role of this expert had just started. Following, the graphs drawn for all the variables by the experts are presented.

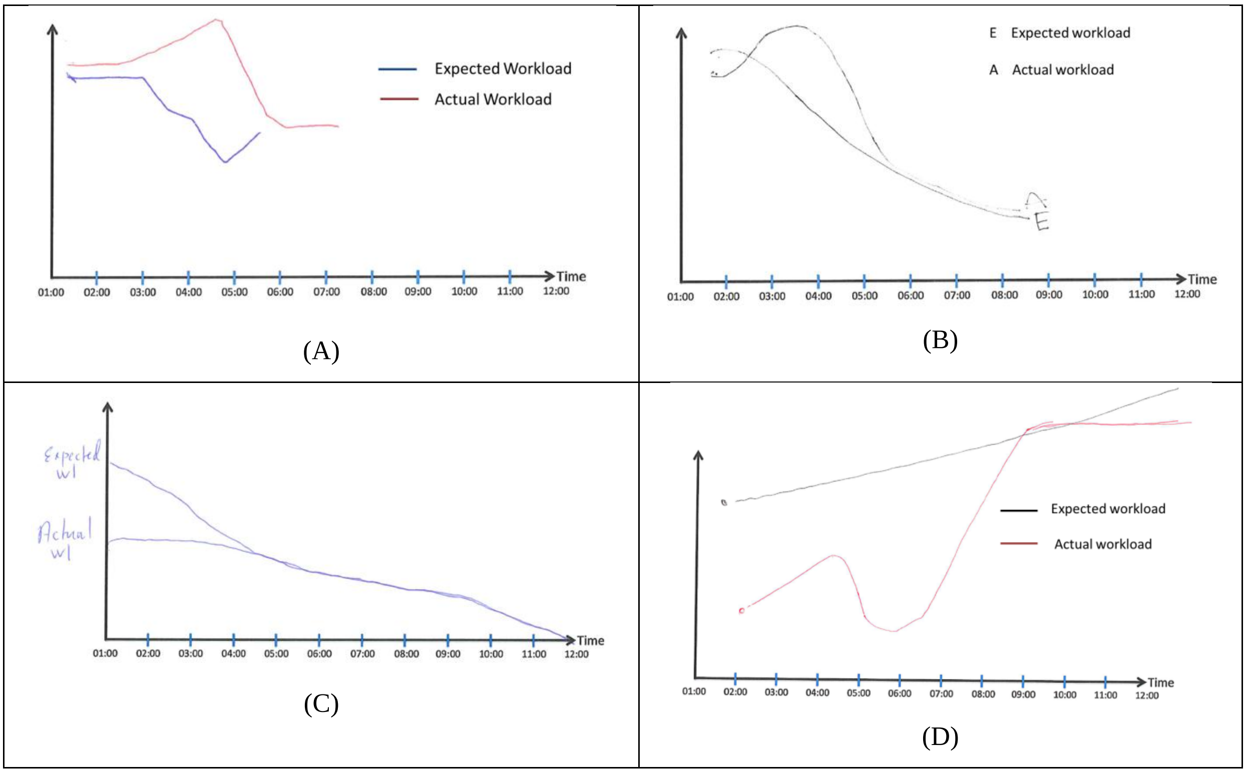


Figure 1 Expected/Actual Workload

The expected workload was initially higher than the actual workload owing to initial delays in the response. Later, the actual workload was higher than the expected workload because unexpected tasks arose. The workload was high initially and decreased along with the response and recovery activities. The graph D is different from the others since, as explained before, the role of the fourth expert was different. He started responding to the tragedy around 3 am. Around 6 am, he arrived to the evacuee center to help to evacuate people and provide support. Therefore, the B graph best represents the reference mode for expected/actual workload.

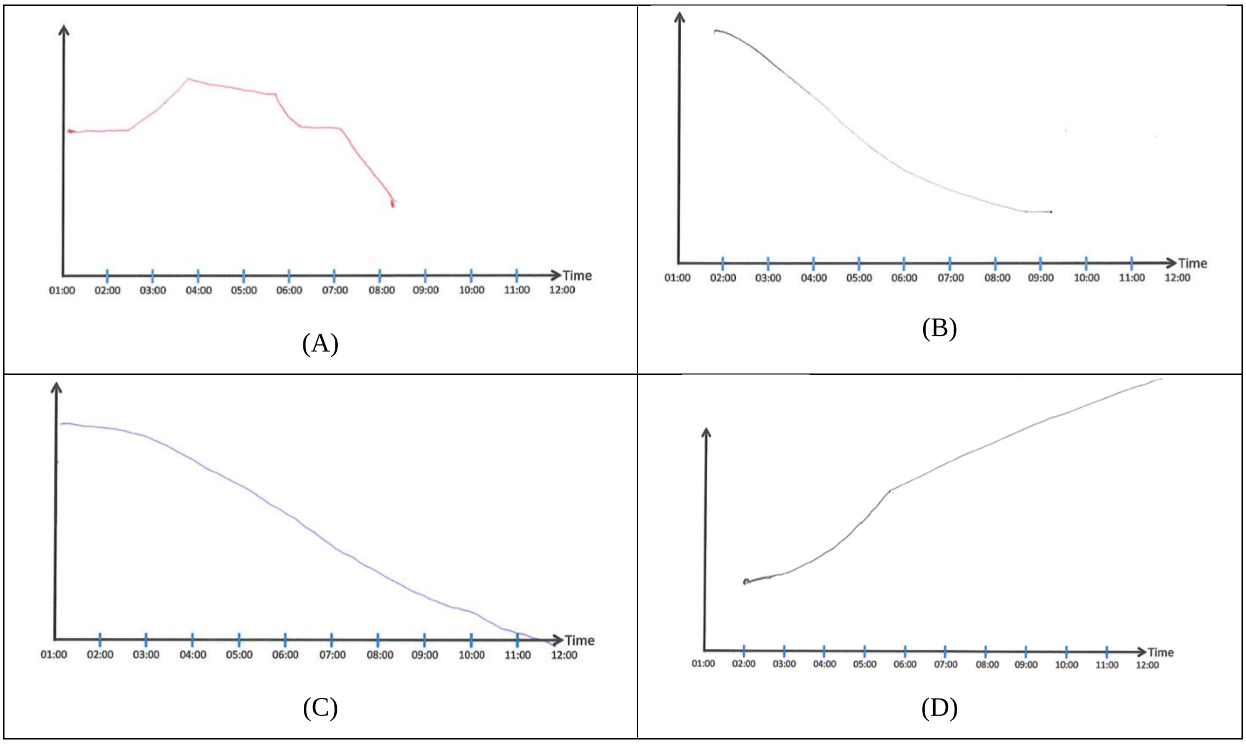


Figure 2 Cognitive Load

When the emergency occurred the disaster situation was new for the responders and therefore, they had to process a high amount of information and interactions to understand the situation and make decisions about how to respond. Once the situation was understood and the disaster started to recover, the cognitive load decreased eventually. Similarly to the previous case, the fourth graph represents a different situation since the role of this expert was different. Among graph A, B, C, we choose graph A as reference mode as the need to process information and task actually increases at the beginning, therefore, the cognitive load should not peak at the beginning, rather, increase for a while and then decrease.

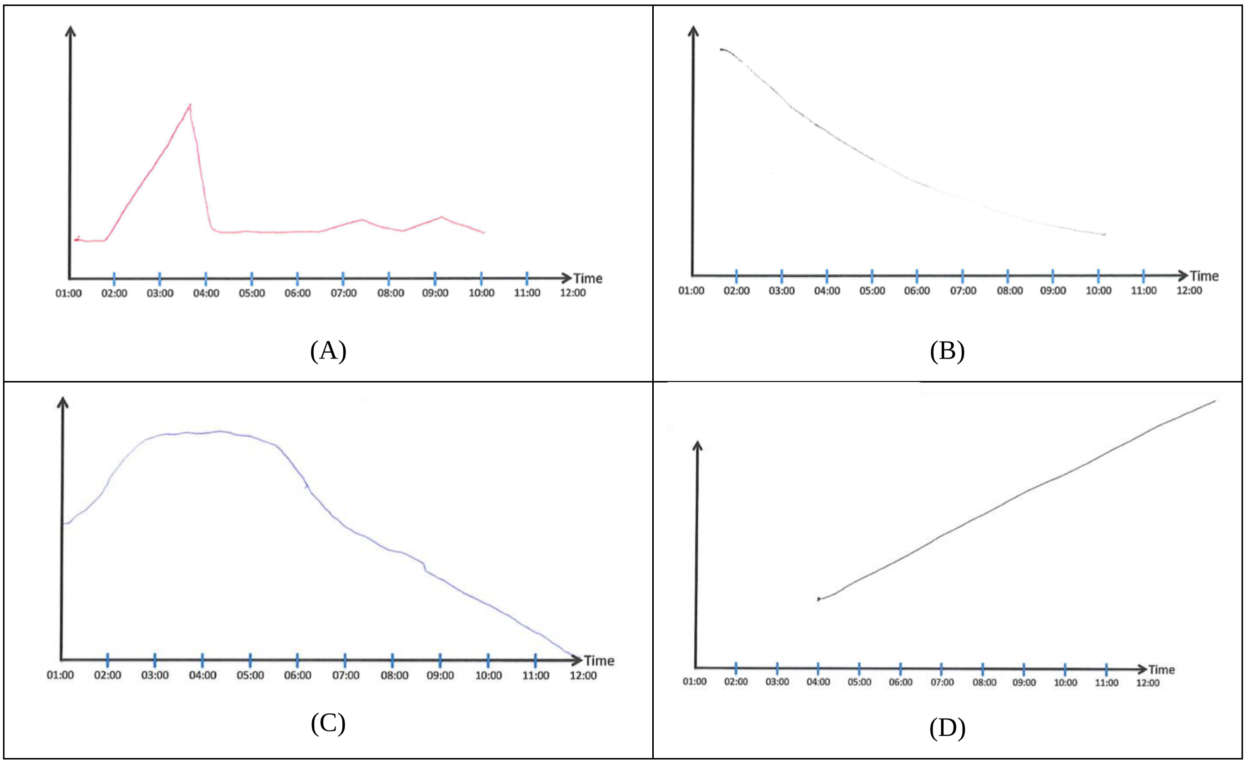
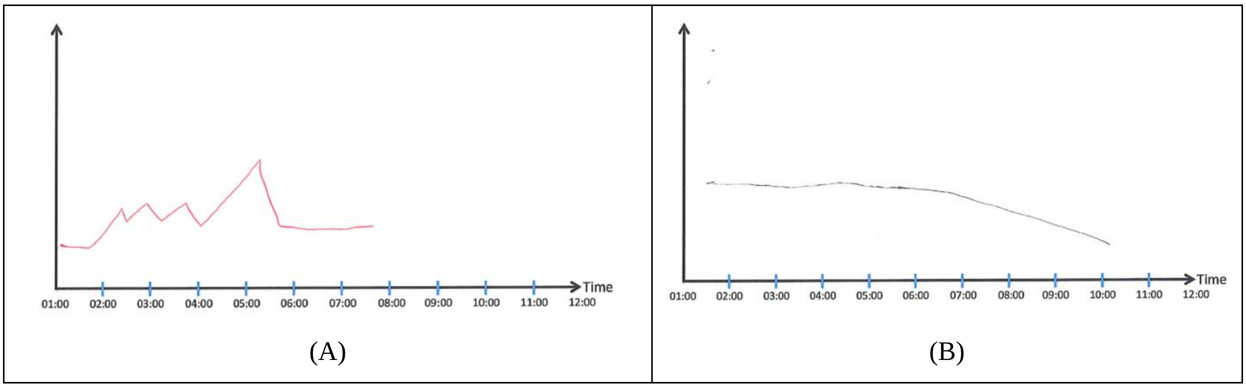


Figure 3 Improvisations

In the beginning, as the situation was not previously predicted, several improvisations were needed in order to respond effectively. In the first hour, the number of improvisations was high and eventually it decreased. As in the previous cases, the graph D has a different behavior. In this case, the number of improvisations that need to carry out increases eventually because they have more and more affected people to manage and provide support. Graph A and C are similar, both expressing the idea that improvisations increases at the beginning and then gradually reduce. We pick Graph A as reference mode.



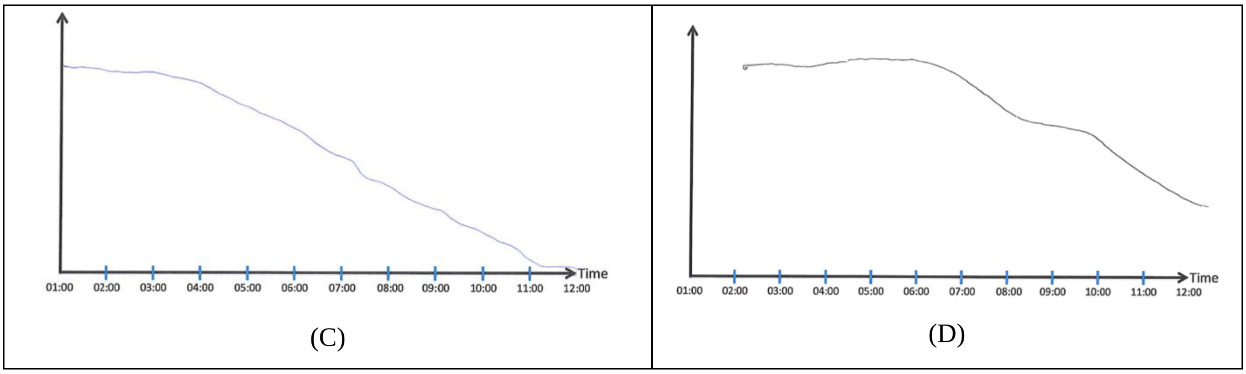


Figure 4 Mismatch of understanding

Although coordination activities are developed within the preparation period to improve the coordination among responders, stakeholders argued that initially they had some trouble to understand communications from other the teams but eventually, after several hours working together, the quality of the mutual communication improved. Accordingly, we choose graph A as the reference mode for the mismatch of understanding.

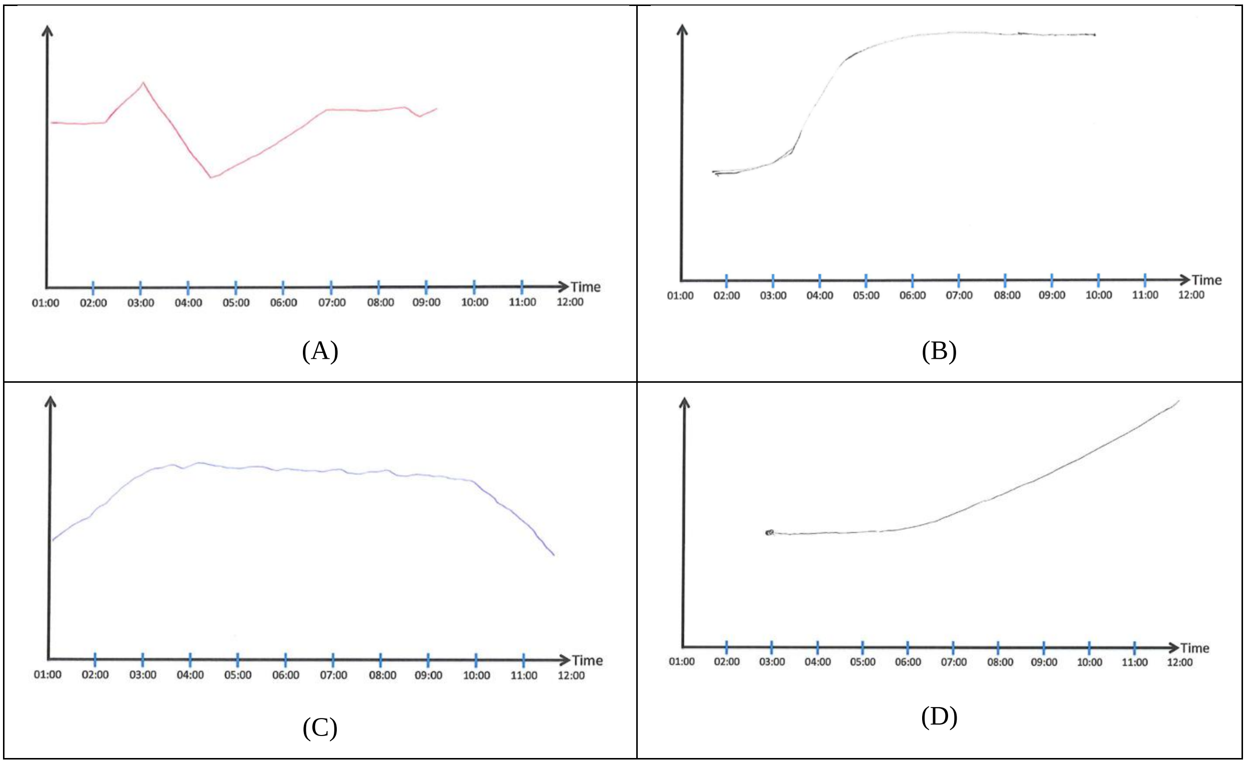


Figure 5 Quality of the decisions

In the beginning the stress situation was high and therefore, stakeholders were not able to make the most appropriate decisions. They experienced several trial and error situations. After some time, the stress situation decreased and as a result, the quality of the decisions increased.

In order to develop the simulation model, the relationships between the main variables are required. We asked the responders for the relationship of the following two pair of variables: the relationship between quality of the decisions and cognitive load and the relationship between quality of the decisions and mismatch of understanding.

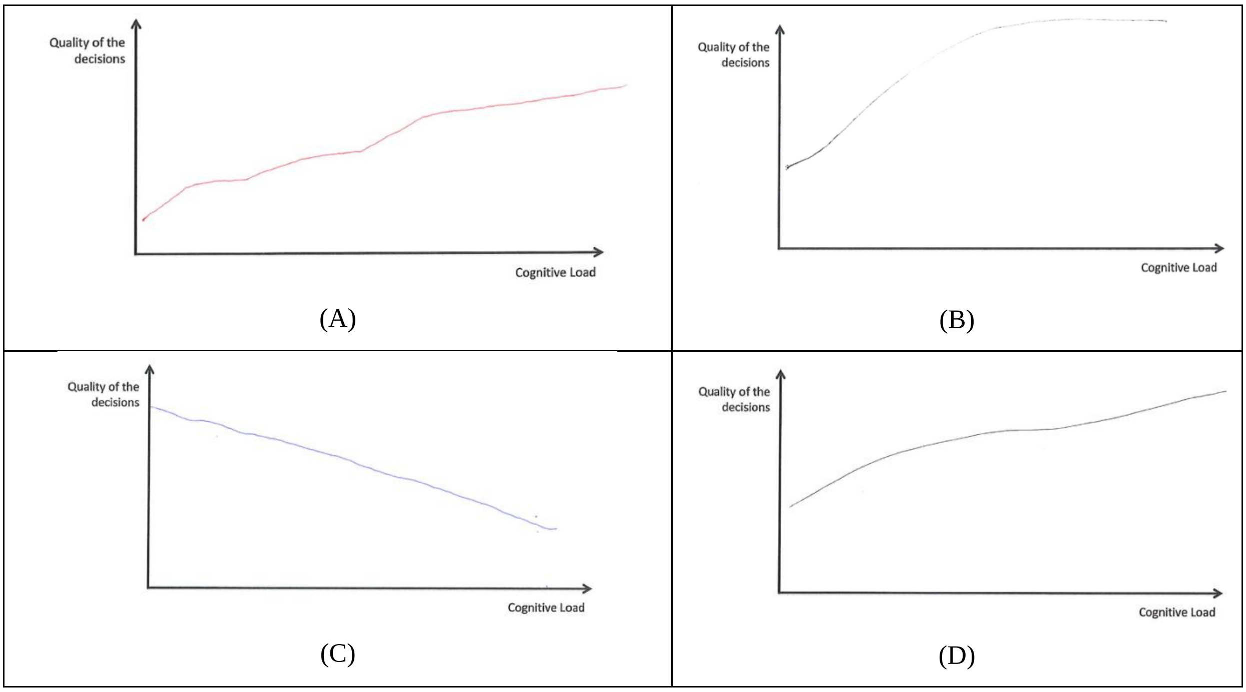


Figure 6 Relationship between quality of the decisions and cognitive load

In this graph the aim was to determine the relationship between the previously defined two variables: quality of the decisions and cognitive load. However, the experts understood this graph differently. Some of them stated, "more thinking leads to better decisions". This means that if responders have more time to think the quality of the decisions will be better since they will be able to analyze all the consequences that a decision may have (graph A, Band D). Having high cognitive load refers that responders have not enough time to properly think about the decision and therefore, the quality of the decisions will be low. On the other hand, the experts number three stated that if the quality of the cognitive load is high then the quality of the decisions will be low and the opposite (see graph C). However, both perspectives lie within the same basic idea: high cognitive load leads to making low quality decisions. Therefore, we choose graph C to represent this reference mode.

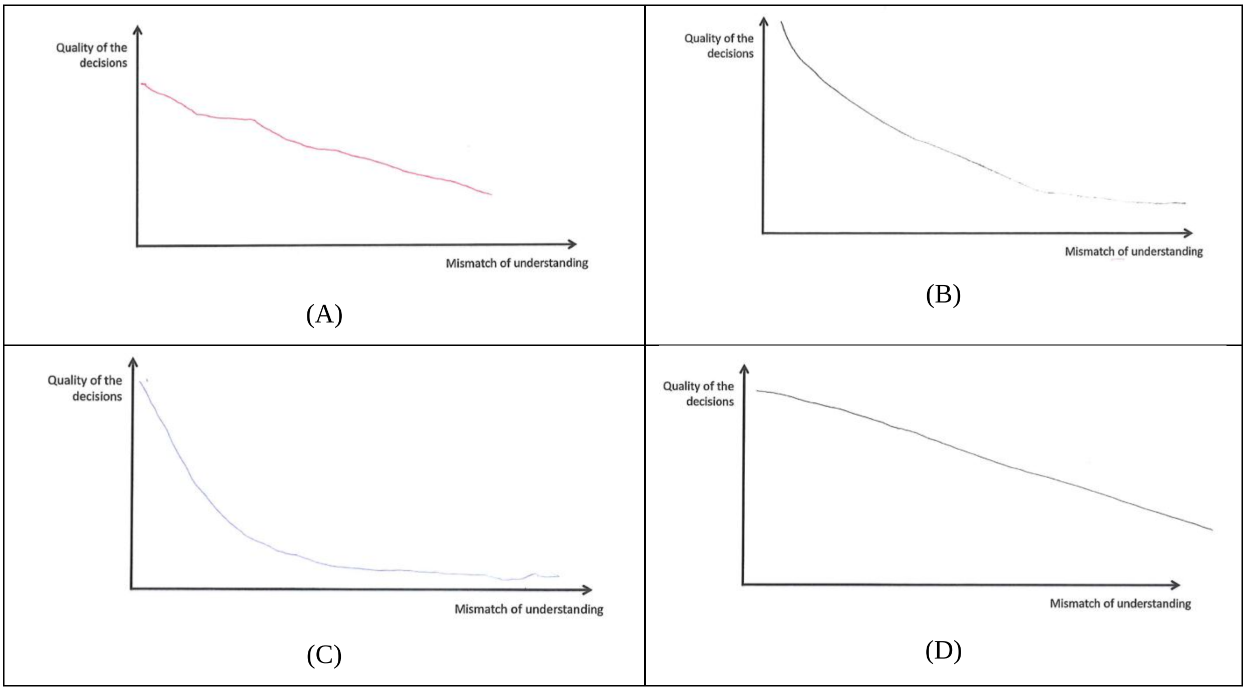


Figure 7 Relationship between quality of the decisions and mismatch of understanding

The target of this graph was to determine the relationship among the previously defined two variables: quality of the decisions and mismatch of understanding. In this case, all the experts agreed on the behavior of this graph.

All the experts considered that the relationship among the quality of the decisions and mismatch of understanding is the opposite relationship. Higher the mismatch of understanding lower the quality of the decisions will be and the opposite. Therefore, all of the four graph express a similar behavior and we can choose either one as reference mode.

THE STRUCTURE OF EMERGENCY MANAGEMENT SYSTEM DYNAMICS MODEL

Dynamic hypothesis

Previous literature (Tu, Wang et al. 2009) and interviews with Hatlestad emergency responders suggest the dynamics hypothesis. Tu et al. proposed that a simple core causal loop diagram with two reinforcing loops and two balancing loops represents the handling of the Palau incident. To confirm that this hypothesis also would apply for the case of the management of the Hatlestad emergency, we interviewed the Hatlesstad emergency responders. .

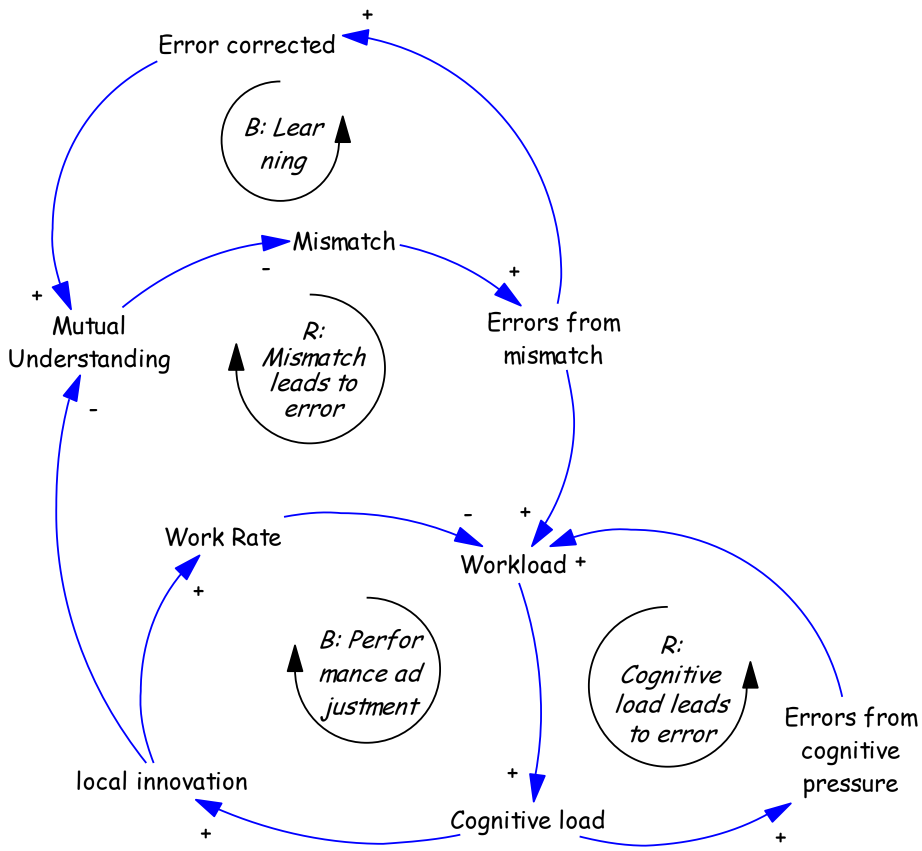


Figure 8 Causal loop diagram

An important balancing loop is the “Performance adjustment” loop. When the emergency happens additional workload is generated. The cognitive load of persons handling the emergency will increase accordingly. To reduce the cognitive load, officers will initiate local innovations and the work rate will improve. Thus, the workload decreases.

Two reinforcing loops relate to errors, viz. “Mismatch leads to error” and “Cognitive load leads to error”. Errors can occur owing to the cognitive load and the mismatch in communication. Upon the onset of the emergency happens the cognitive load increases and errors will be generated from cognitive pressure. This will further increase the workload. Under heavy workload and high cognitive load, local innovations take place. Normal working routines are thus disrupted and mutual understanding drops, leading to mismatch among team members. This will breed errors from mismatch, which eventually cause additional workload.

A balancing loop, “Learning”, expresses that error correction gradually increases mutual understanding and reduces mismatch, which leads to fewer errors.

Model sectors

Workload and cognitive load

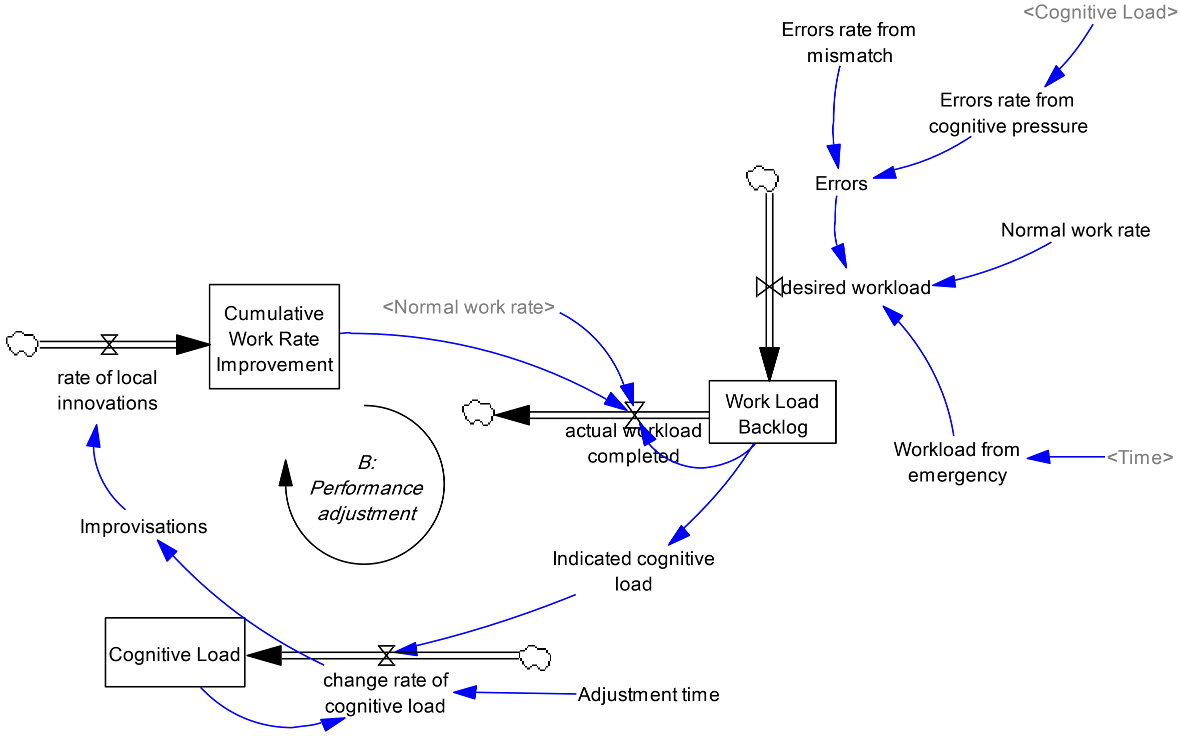


Figure 9 Model structure—Workload and cognitive load

The variable '*expected workload*' represents the quantity of tasks to be completed during the emergency, which is composed of '*expected workload from emergency*', '*normal work rate*' and '*errors*' (referring to the added workload caused by errors).

$$expected\ workload = expected\ workload\ from\ emergency + normal\ work\ rate\ + errors$$

'*Workload Backlog'* is drained by '*actual work completed'*. When '*expected workload'* is higher than '*actual work completed'*, the backlog of workload increases. '*Workload Backlog'* generates cognitive load. In accordance with the disequilibrium–experimenting–emergence paradigm the responders react to the disruption induced by a high cognitive workload with improvisations. Over time (after a learning process, see below) the improvisations improve the work rate. The variable '*actual work completed*' increases and this reduces '*Workload Backlog'* and '*Cognitive Load'*. This part represents the balancing loop "Performance adjustment".

The variable '*expected workload from emergency'* represents the additional workload caused by the disaster. In the Hatlestad slide case, according to the interviews, the additional workload from emergency is high at the beginning and gradually decreases to zero as more and more routine tasks replace the emergent task from disaster conditions. The variable '*errors'*, representing the added workload from errors, include errors generated by mismatch and errors generated by cognitive pressure.

Mismatch and Errors

When local innovations are implemented, the "Mutual understanding" decreases. The "Mutual understanding" is measured as 0-1 scale, representing no mutual understanding/full mutual understanding. Mismatch will be created when "Mutual understanding" is less than 1.

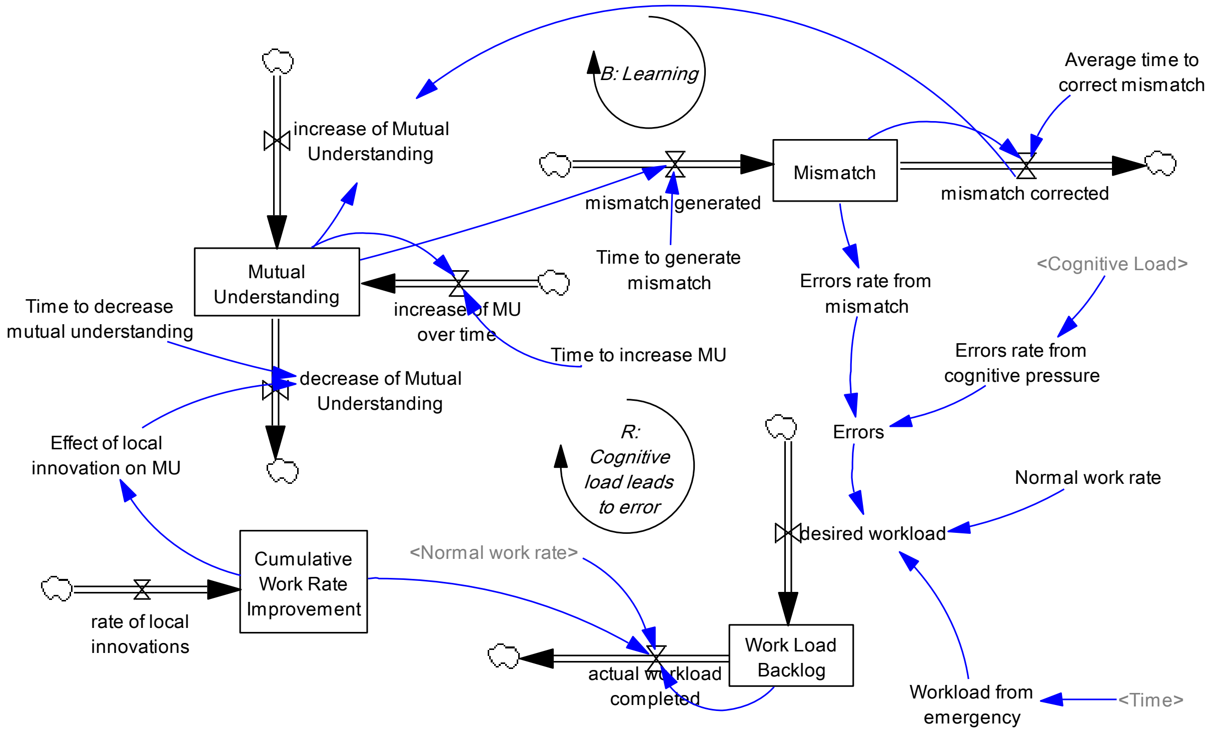


Figure 10 Model structure—Mutual understanding and mismatch

Mismatch will be corrected over time. During the process of correcting mismatch, officers will communicate with each other, and thus, the "Mutual understanding" will increase. When "Mutual understanding" rises, the "Mismatch" will be lower than it otherwise will be. This represents the balancing loop "learning".

SIMULATION AND FEEDBACK ANALYSIS

Stage 1: From the 1st to the 90th minute.

Figure 11 illustrates variables related to workload. The blue line with number 1 is "Normal workload", which remains consistently at 10 task/minute. When emergency happens, additional tasks occur, which are represented by "Workload from emergency". This variable is an exogenous input starting at 6 task/minute (60% of normal workload) and gradually reducing to 0. This corresponds to the Hatlestad Slide case situation. For other cases, "Workload from emergency" could be set differently.

The green line (3) represents the "desired workload". When the emergency happens, "Workload from emergency" leads to unusually high "Desired workload". Facing high "desired workload", responders will implement local innovations to improve work efficiency and try to complete the work. However, there is a delay in the implementation of the local innovation, therefore, the '*actual workload completed'* follows the growth of '*Expected workload*' with a gap. When '*actual work rate*' is lower than '*Expected workload*' workload required, the workload backlog increases and so does the "Cognitive Load". Higher cognitive load leads to more errors from cognitive load. At the same time, new local innovations reduce mutual understanding and cause errors from mismatch. The "Desired workload" increases in the first hour mainly because the additional tasks to correct the errors (see figure 12). The workload from emergency gradually decreases and the "actual workload completed" increases due to the local innovations.

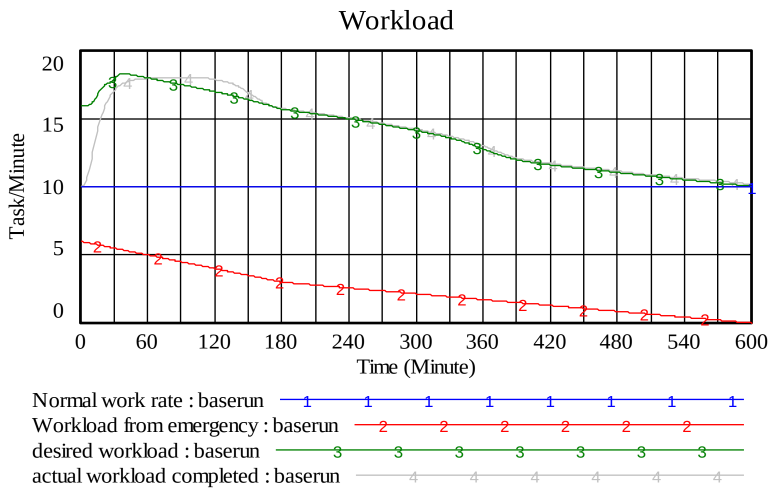
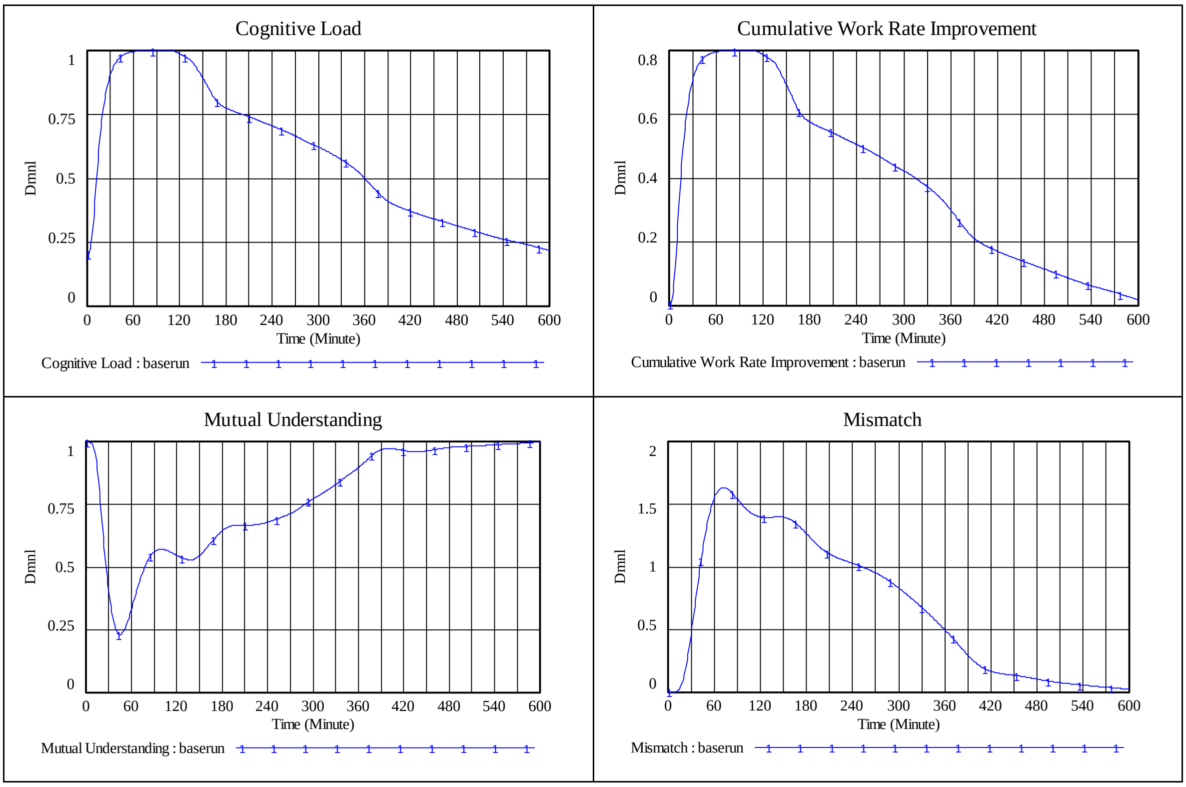


Figure 11 Simulation result—workload

In figure 12, we can see that the “Cognitive load” peaks at the point that “actual work completed” cross over the “desired workload”. When actual work rate is lower than workload required, the workload backlog increases and so does the “cognitive load”. When “Cognitive load” increases, the local innovations are initiated and work rate is improved. The “Cumulative work rate improvement” behaves similarly as “Cognitive load”. When innovations are implemented, mutual understanding is reduced and mismatch is increased, which causes high “errors rate from mismatch”. Errors rate means the percentage of errors in the total work completed.



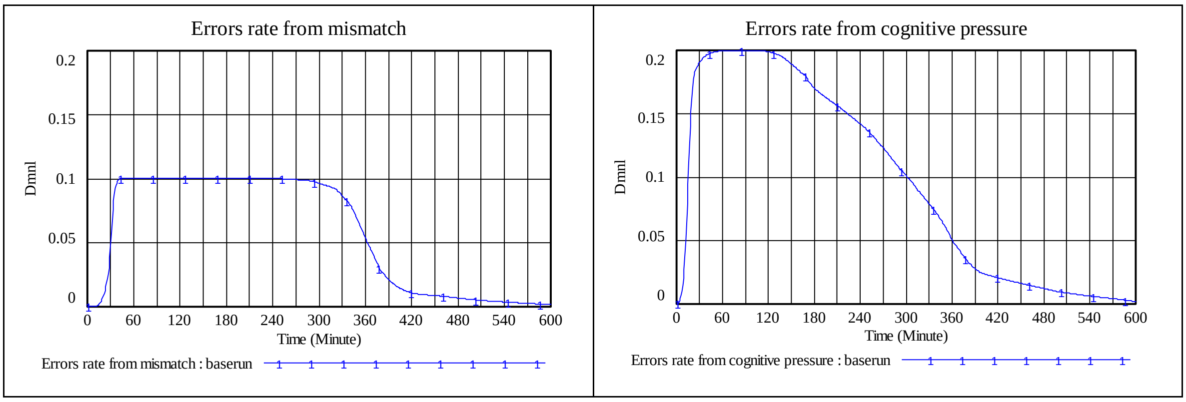


Figure 12 Simulation result

The “Errors rate from cognitive pressure” is highly related to “cognitive load”. Therefore, “Errors rate from cognitive pressure” behaves in the same way as “cognitive load”. Regarding “Errors rate from mismatch”, when “cognitive load” is high, local innovation happens and this reduces mutual understanding. As a result, the errors generated by mismatch increases sharply at the beginning. A combination of the above-mentioned two types of errors, “errors” increase sharply at the beginning, perk around 90 minutes and then starts to decrease. These errors add additional tasks to “desired workload”. Therefore, the “desired workload” increases in the first hour.

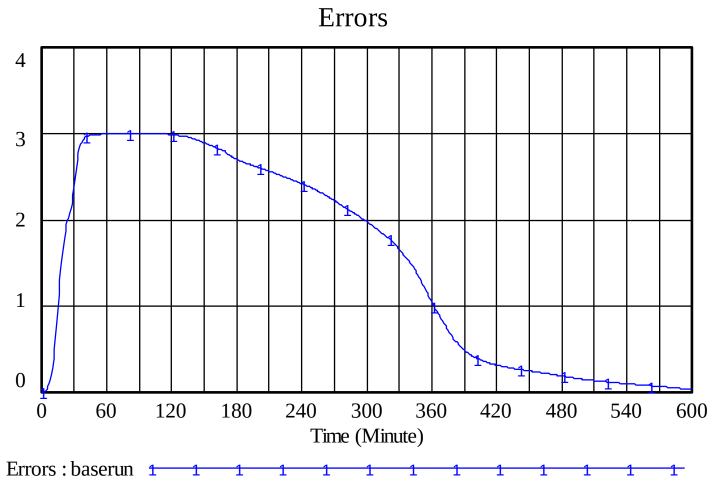


Figure 13 Simulation result—Errors

The system behavior in the first one and a half hours is dominated by the two reinforcing loops “Cognitive load leads to errors” and “Mismatch leads to errors”. At the same time, the balancing loop performance adjustment is working, the “actual work completed” is increasing. However, as there is delay in the system, the “actually work completed” is still lower than the “desired workload”.

Stage 2: From the 90th minute to the end.

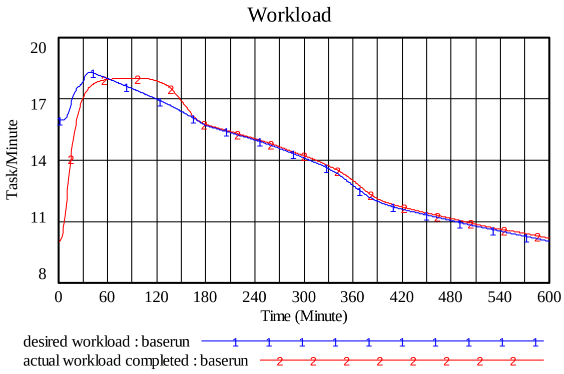
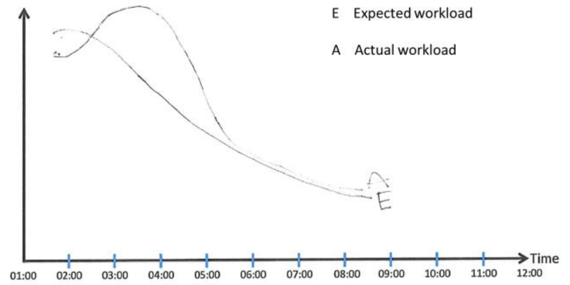
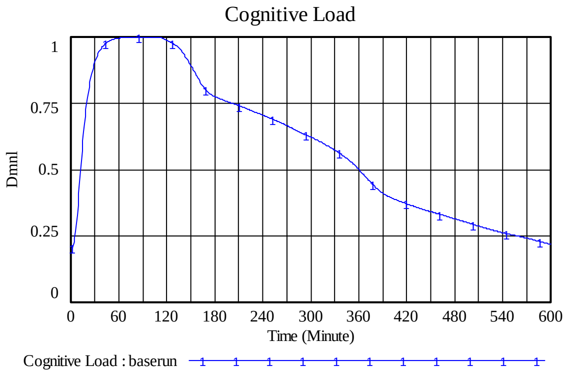
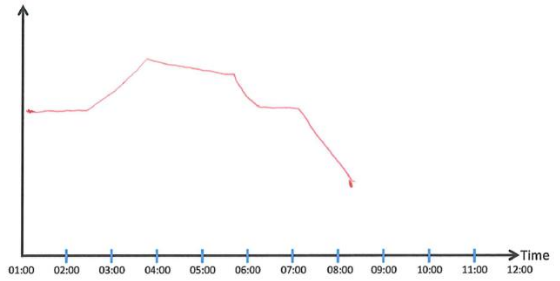
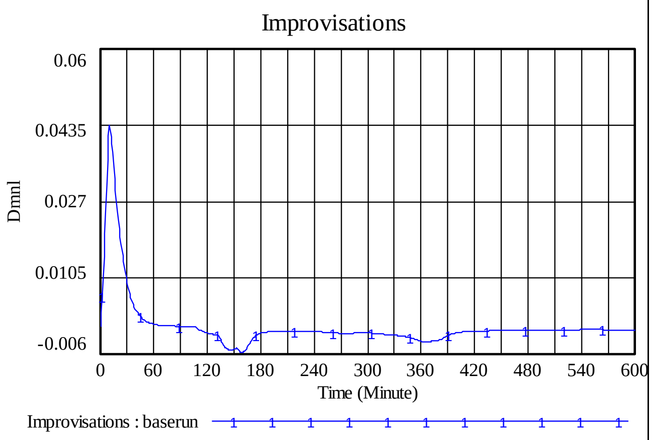
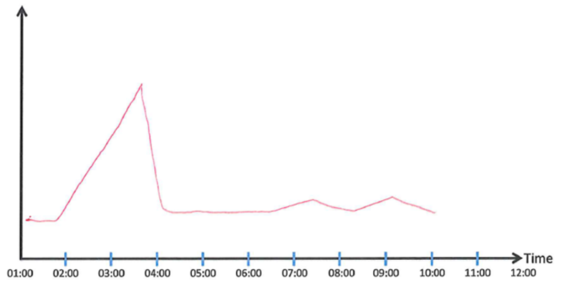
Over time, the local innovations are implemented and the cumulatively the work rate is improved. The “actual work completed” increases to the level that is higher than the “desired workload”. This leads to a reduction of “workload backlog”. Therefore, after the first 90 minutes, the major loop that dominates the system behavior is the balancing loop—performance adjustment. The “cognitive load” starts to reduce and as a result, the errors generated from cognitive load are reduced. As responders work together, mutual understanding increases and errors from mismatch also reduces. Moreover, the correction of the mismatch leads to learning, which also raises mutual understanding. Fewer errors are generated. As a result, the “desired workload” is lower than it otherwise could have been.

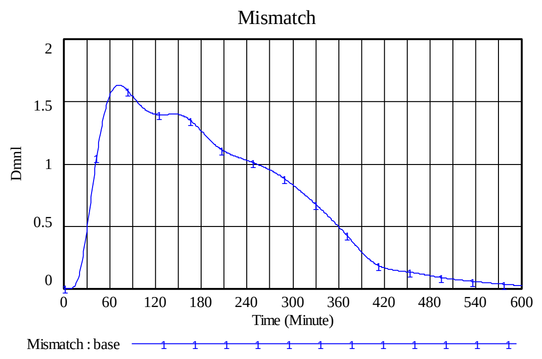
In the stage, the balancing loop performance adjustment’s effect become obvious and then the system switch from a vicious loop (more errors generate more work, which add to workload and generate even errors) into a virtuous loop (less workload backlog, reduces the errors, which reduces the workload further). The other balancing loop of learning is also dominant during this stage

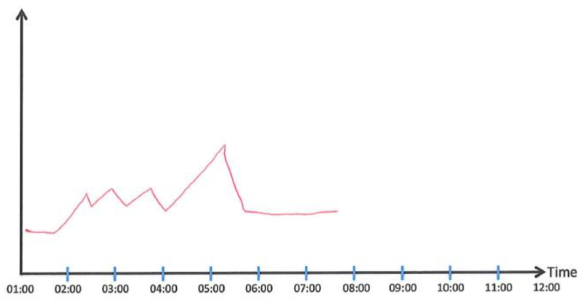
COMPARISON WITH REFERENCE BEHAVIOR

System dynamics model's robustness lies in the correctness of the model structure. The model composes four causal loop diagrams that are in accordance with the interview of the personnel handling the emergency. We compare the reference mode provided by the informants with the model behavior.

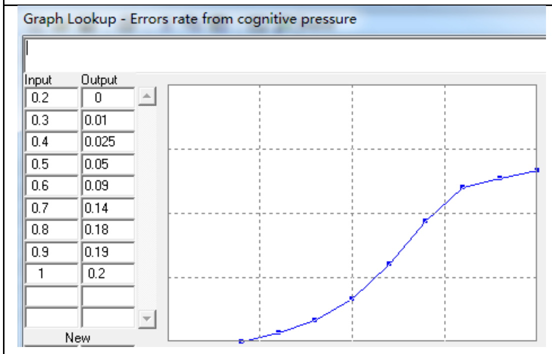
Table 2 Comparison of model behavior with reference mode

Desired workload and actual workload completed	
	
Cognitive Load	
	
Improvisations	
	
Mismatch of Understanding	

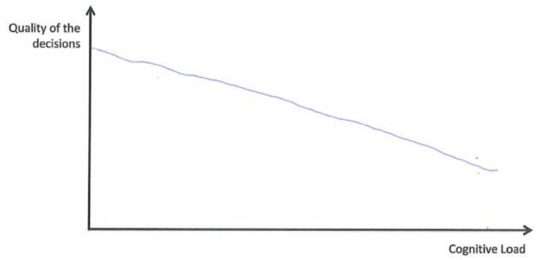




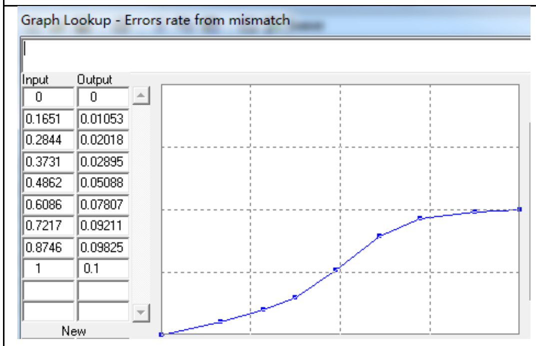
Relationship between quality of the decisions and cognitive load



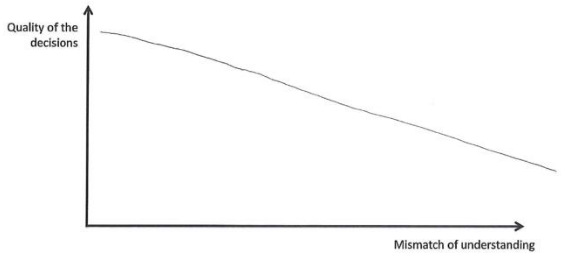
Error rate from cognitive load



Relationship between quality of the decisions and mismatch of understanding



Error rate from mismatch



The next two sections concern extreme tests and sensitivity analysis of the model.

EXTREME TESTS

In this section, we conduct a set of extreme test to validate the model. Below is a table summarizing the extreme tests with the detailed information on parameter change.

Simulation name	Base run parameter setting	Extreme test parameter setting
Ext1 No emergency	WITH LOOKUP(time, ([(0,0)-(800,20)],(0,6),(180,3),(600,0)))	Workload from emergency (=0)
Extreme long time to adjust cognitive load	Adjustment time=10 minutes	Adjustment time=200 minutes
Extreme long time to time to decrease mutual understanding	Time to decrease mutual understanding=3 minutes	Time to decrease mutual understanding=200 minutes
Extreme long time to generate	Time to generate mismatch=10	Time to generate mismatch =200

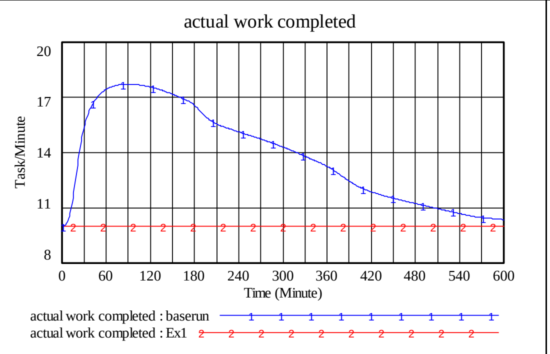
mismatch	minutes	minutes

Ext1 No emergency

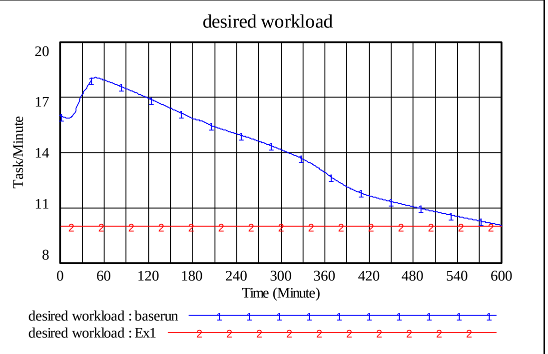
Variable change: Workload from emergency (=0)

Expected result: Without any workload from emergency, the system will be in equilibrium condition. No variable will change.

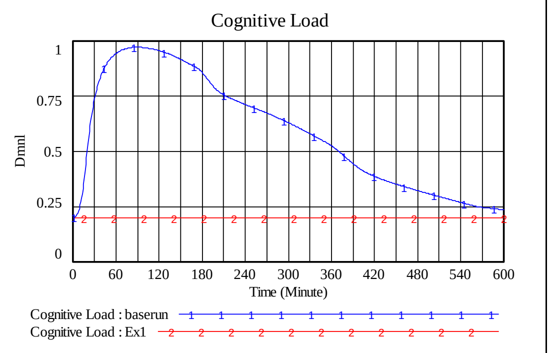
actual work completed



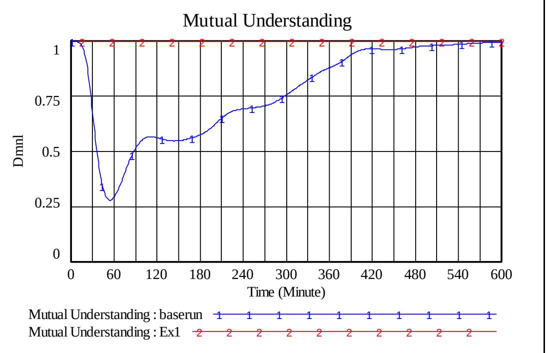
desired workload



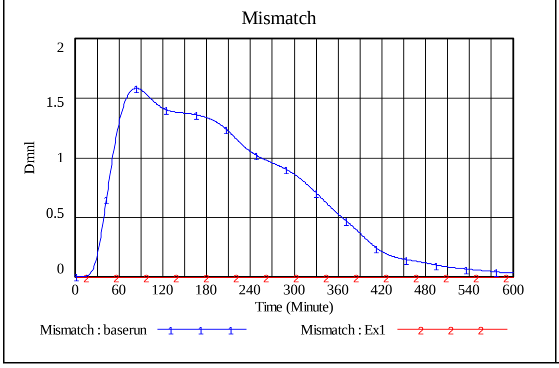
Cognitive Load



Mutual Understanding



Mismatch



Errors

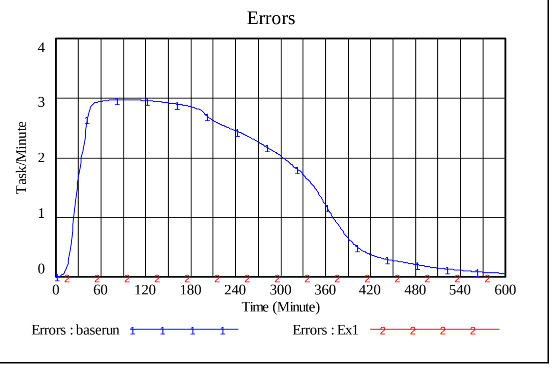


Figure 14 Ext1 No emergency. Model behavior:

(Base: Blue line, with number 1; Extreme test: Red line, with number 2)

Result: Model simulated results fit the expectation. Extreme test passed.

Extreme long time to adjust cognitive load

Variable change: Adjustment time (=200)

Expected result: Cognitive load changes slowly, therefore, the peak will come later. Similarly, other variables also peak later.

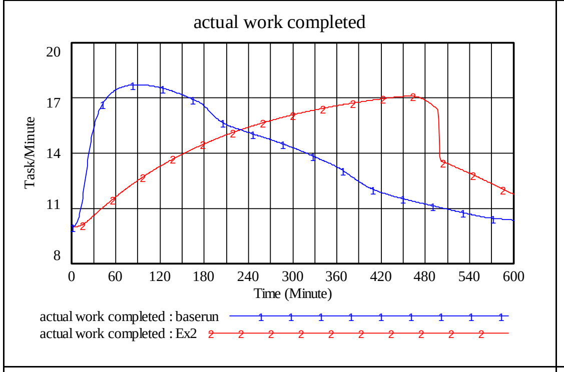
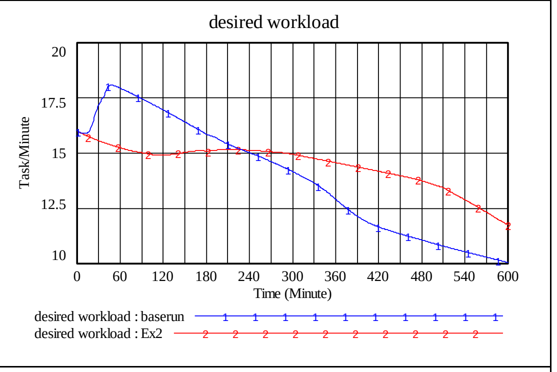
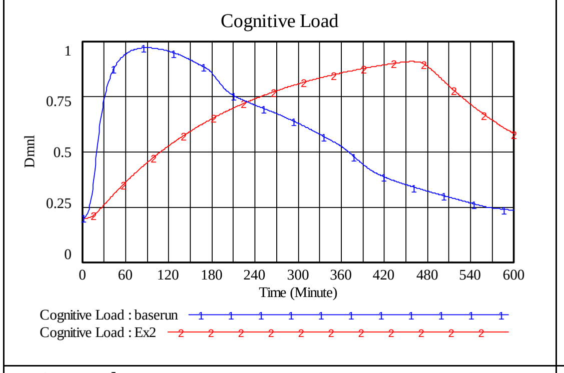
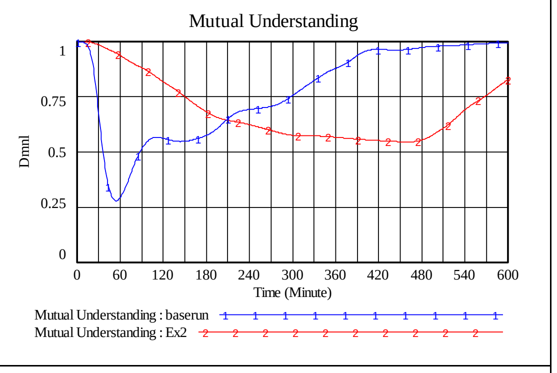
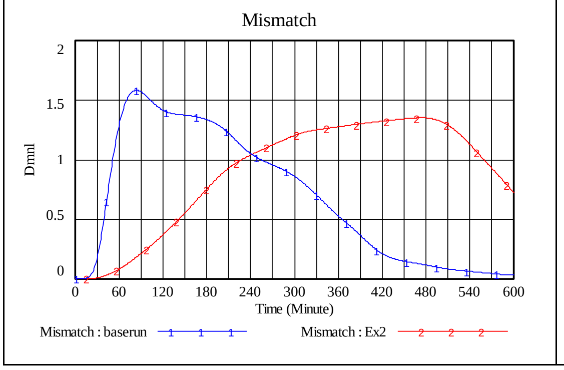
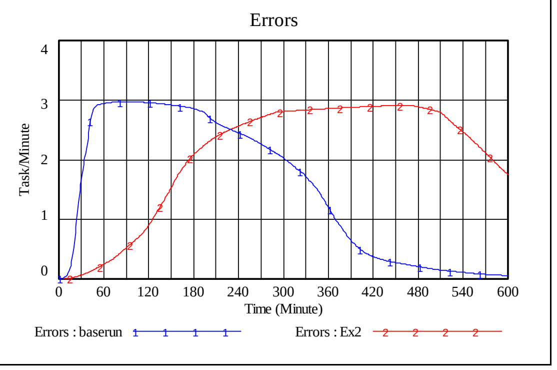
actual work completed	desired workload
	
Cognitive Load	Mutual Understanding
	
Mismatch	Errors
	

Figure 15 Ext2 Extreme long time to adjust cognitive load. Model behavior: (Base: Blue line, with number 1; Extreme test: Red line, with number 2)

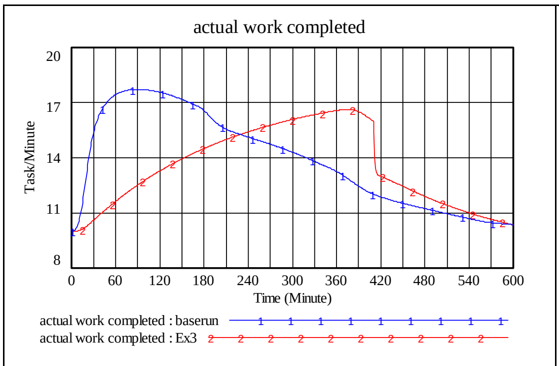
Result: The desired workload doesn't increase at the beginning. This is because in the baserun, the desired workload increases at the beginning because the increase of errors. In the extreme test, errors doesn't go up quickly, therefore, the desired workload doesn't increase. Extreme test passed.

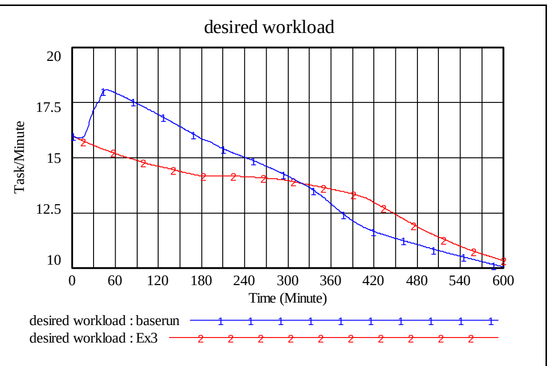
Extreme long time to time to decrease mutual understanding

Variable change: time to decrease mutual understanding (=200)

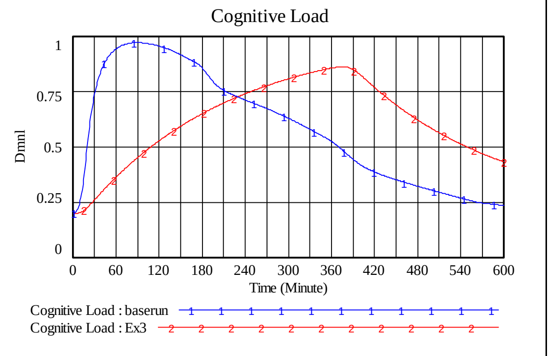
Expected result:: This variable means how fast mutual understanding are reduced when local innovations are implemented. In the baserun, when local innovations happen, in 3 minutes, mutual understanding will be reduced. In this extreme text, the mutual understanding will be reduced in 200 minutes. Therefore, "Mutual understanding" and "Mismatch" will not change as much as baserun. "Errors" will be smaller and thus the "desired workload" will not increase. "Cognitive load" will not be as high and "actual work completed" will not increase as quickly

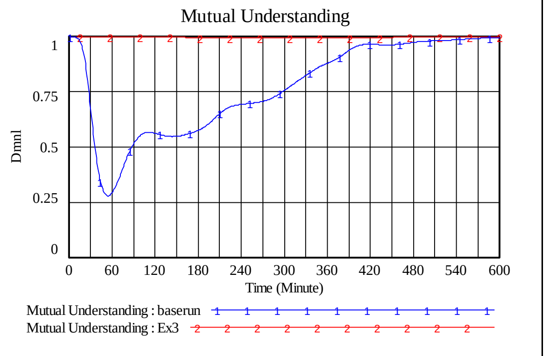
actual work completed	desired workload





Cognitive Load	Mutual Understanding





Mismatch	Errors

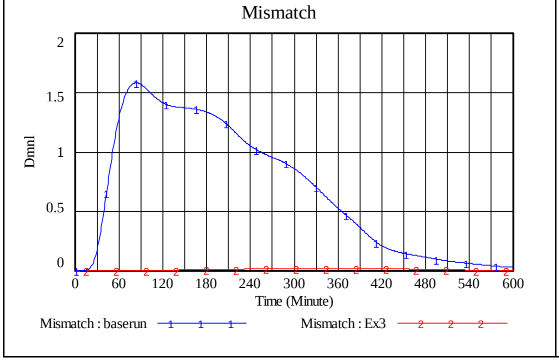


Figure 16 Ext3 Extreme long time to time to decrease mutual understanding. Model behavior: (Base: Blue line, with number 1; Extreme test: Red line, with number 2)

Result: Model stimulated results fit the expectation. Extreme test passed

Extreme long time to generate mismatch

Variable change: time to generate mismatch (=200)

Expected result: "Mismatch" changes slower than baserun and thus "Errors" will be less. Less learning happens and "mutual understanding" will be less than baserun. With less "Errors", the "desired workload" will not increase. The "actual work completed" will be slower than the baserun

actual work completed	desired workload

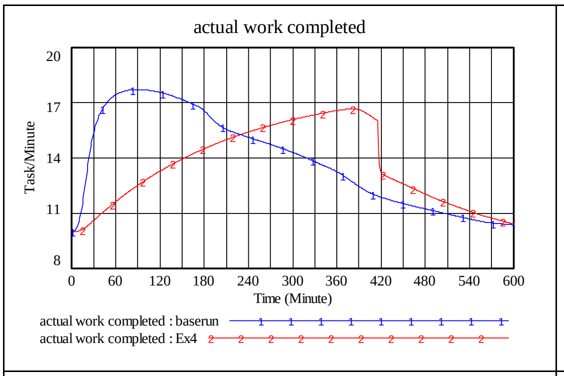
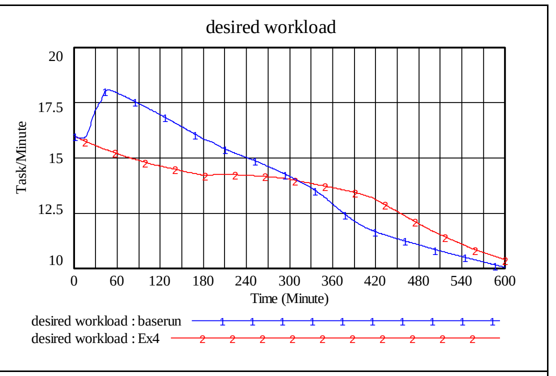
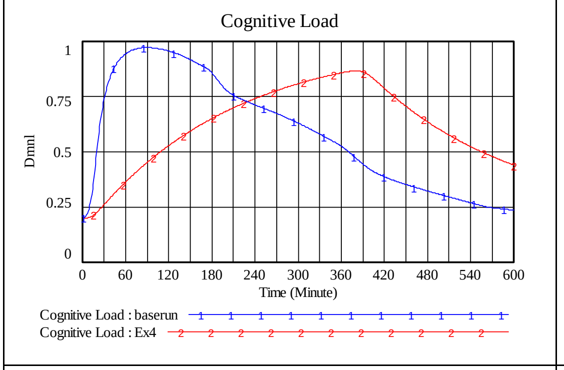
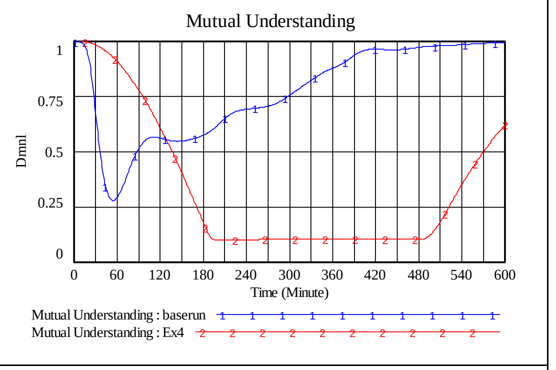
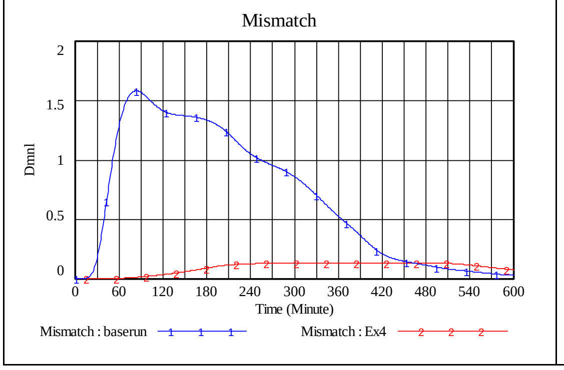
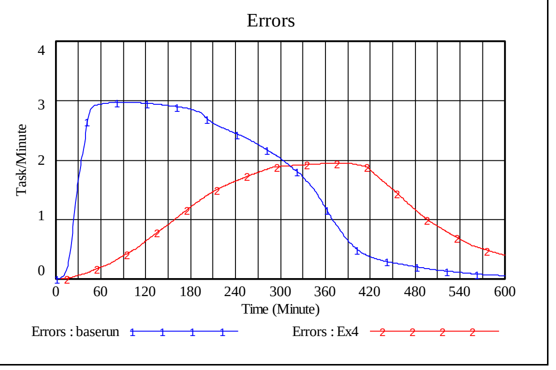
	
Cognitive Load	Mutual Understanding
	
Mismatch	Errors
	

Figure 17 Ext4 Extreme long time to generate mismatch.

Model behavior: (Base: Blue line, with number 1; Extreme test: Red line, with number 2)

Result: Model stimulated results fit the expectation. Extreme test passed.

SENSITIVITY TESTS

In the section below, we conduct sensitivity tests to validate the model. Two kinds of sensitivity tests are performed. For the constants, we will test model behavior of a certain range of the constants. The range is decided based on the rational setting of the constant. For lookup functions, we will test the different shape of the look up, convex, linear and concave.

ST1 Time to decrease mutual understanding

Base run: Time to decrease mutual understanding = 3

Sensitivity test: Range: 1-5; Distribution: Random Uniform; Runs: 200;

Simulation behavior:

actual work completed	desired workload

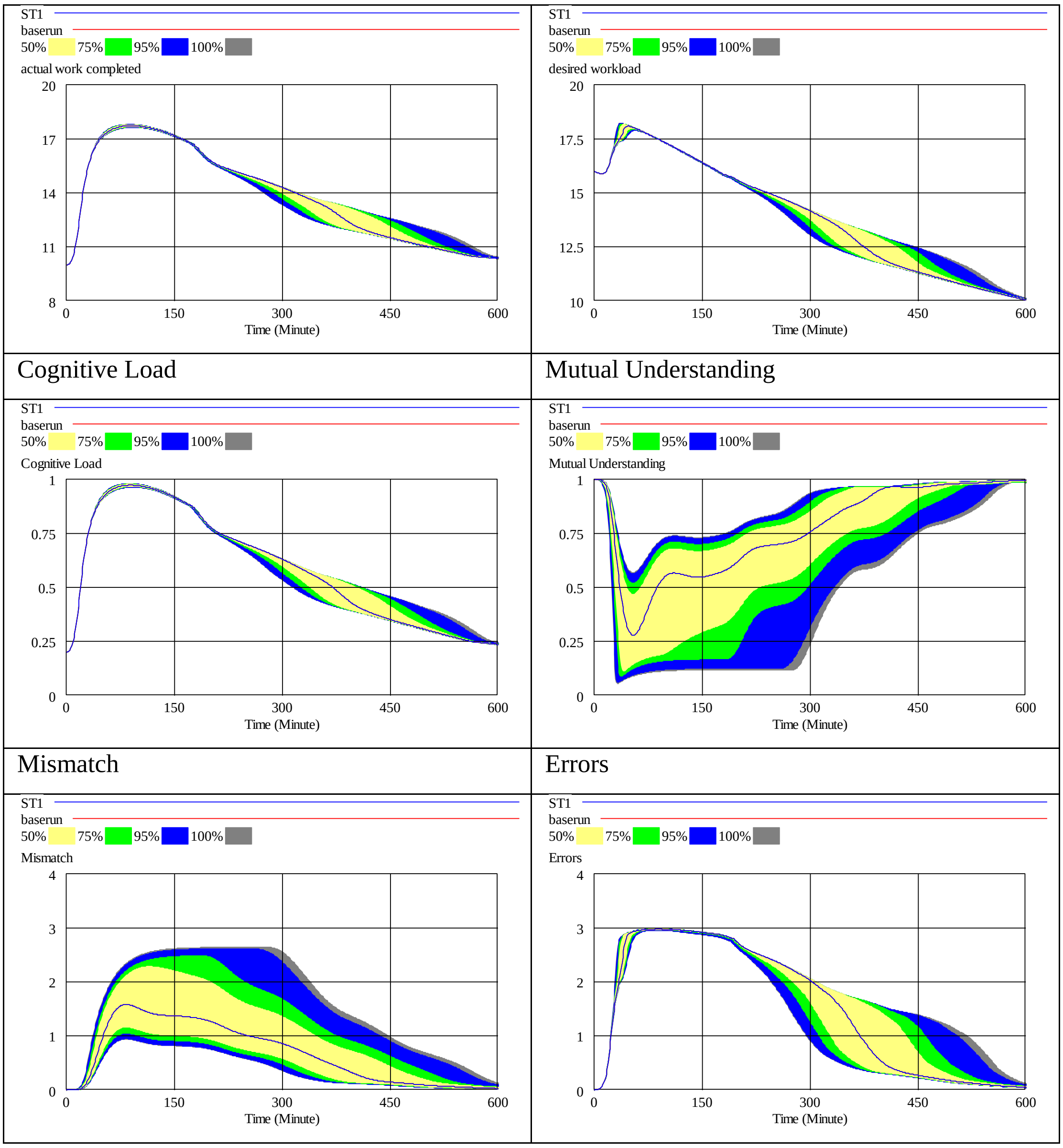


Figure 18 ST1: Time to decrease mutual understanding

Analysis:

The “actual work completed” and the “desired workload” do not change much. However, “Mutual Understanding” and “Mismatch” are quite sensitive to “time to decrease mutual understanding”. This is because “time to decrease mutual understanding” is directly related to “Mutual Understanding”. When “time to decrease mutual understanding” is long, the “Mutual Understanding” changes slowly, and when “time to decrease mutual understanding” is short, the “Mutual Understanding” changes fast. Yet, the “Mutual Understanding” is within the range of 0-1. As “Mismatch” is highly related to “Mutual Understanding”, “Mismatch” is also sensitive to “time to decrease mutual understanding”. “Errors” are partly caused by mismatch; therefore, “errors” also show some sensitivity.

Numerical sensitivity is observed in “Mutual Understanding”, “Mismatch”, and “Errors”. But no pattern sensitivity found. The numerical sensitivity is reasonable regarding the structure of the system.

Result: Pass

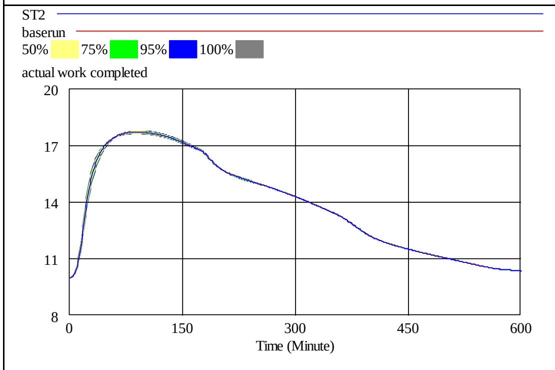
ST2- Adjustment time

Base run: Adjustment time = 10

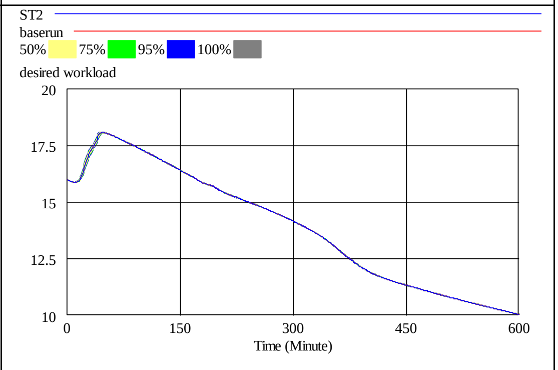
Sensitivity test: Range: 8-12; Distribution: Random Uniform; Runs: 200;

Simulation behavior:

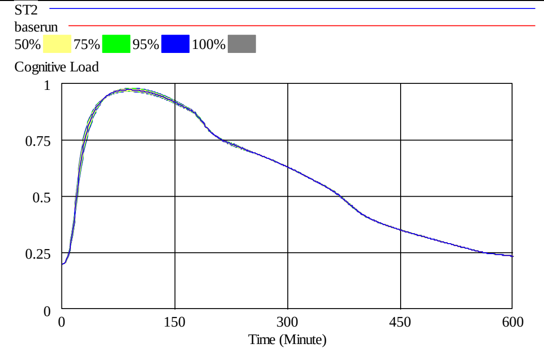
actual work completed



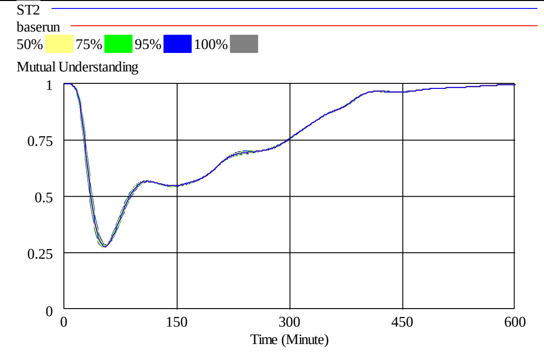
desired workload



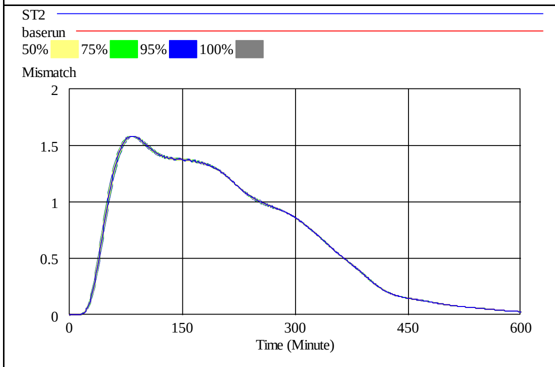
Cognitive Load



Mutual Understanding



Mismatch



Errors

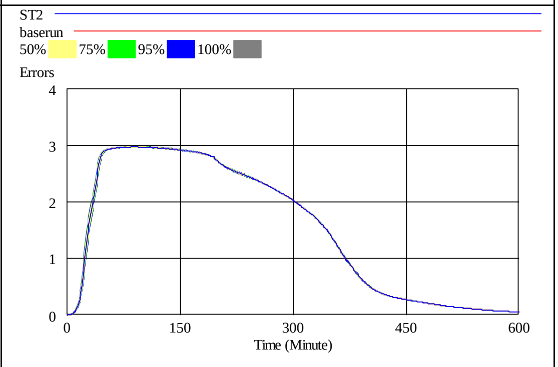


Figure 19 ST2- Adjustment time

Analysis:

The “Adjustment time” is the time to change “cognitive load”. Sensitivity test shows that the “adjustment” has little impact on “actual work completed”, “desired workload”, “Mutual Understanding”, “Mismatch” and “Errors”.

Result: Pass

ST3- Average time to correct mismatch

Base run: Average time to correct mismatch = 30

Sensitivity test: Range: 20-40; Distribution: Random Uniform; Runs: 200;

Simulation behavior:

actual work completed	desired workload

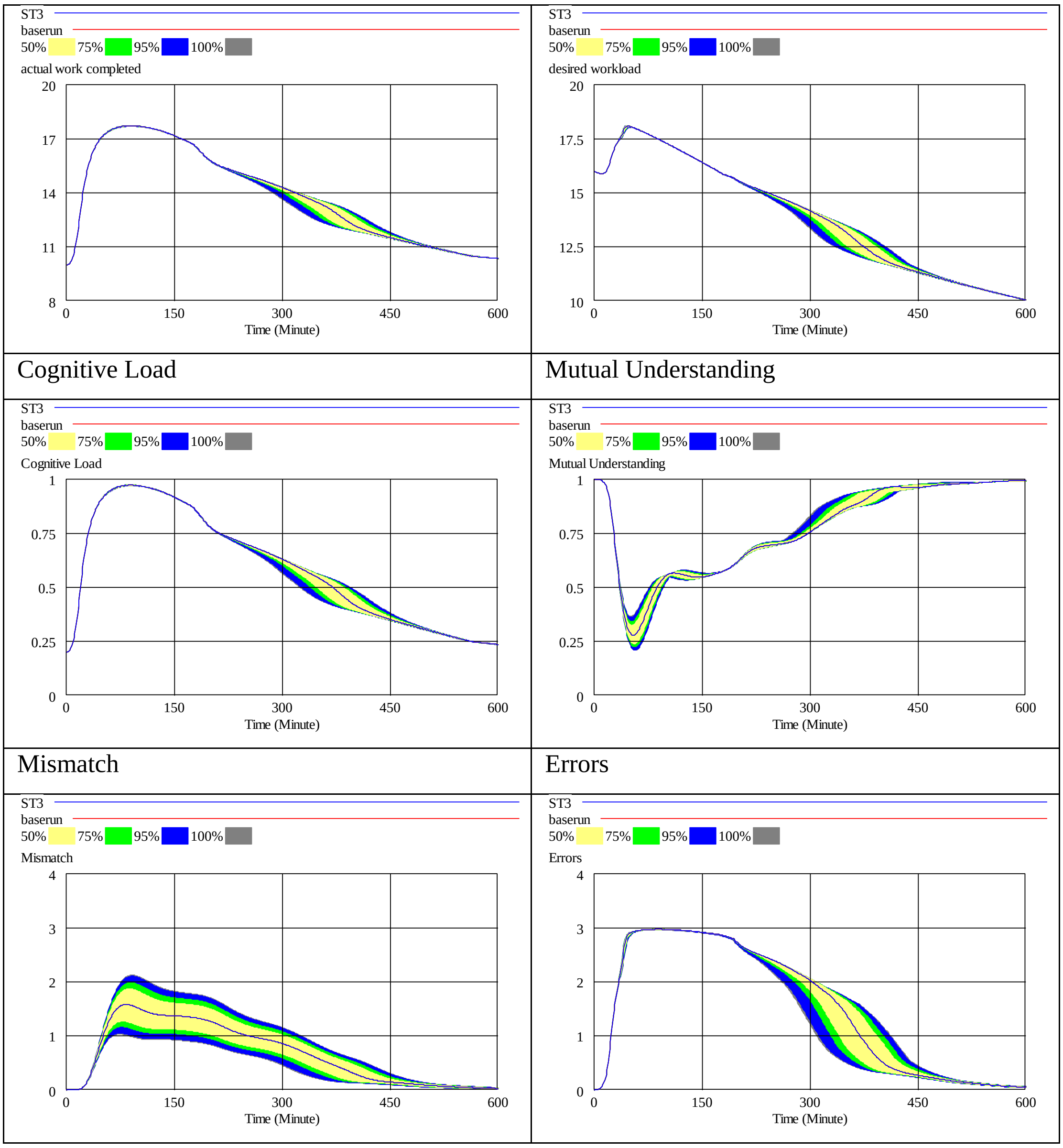


Figure 20 ST3- Time to decrease mutual understanding

Analysis:

The “average time to correct mismatch” has little impact on the “actual work completed”, “desired workload”, “Cognitive load”, and “Mutual Understanding”. However, “Mismatch” is quite sensitive to “average time to correct mismatch”. This is because “time to decrease mutual understanding” is directly related to “Mismatch”. When the “average time to correct mismatch” is long, the “Mismatch” decrease slowly, and when “time to decrease mutual understanding” is short, the “Mismatch” changes fast. However, there is no pattern change of “Mismatch”. As “Errors” are partly caused by mismatch, “errors” also show some sensitivity.

Numerical sensitivity is observed in “Mismatch” and “Errors”. But no pattern sensitivity found. The numerical sensitivity is reasonable regarding the structure of the system.

Result: Pass

ST4- time to generate mismatch

Base run: Time to generate mismatch = 10

Sensitivity test: Range: 5-15; Distribution: Random Uniform; Runs: 200;

Simulation behavior:

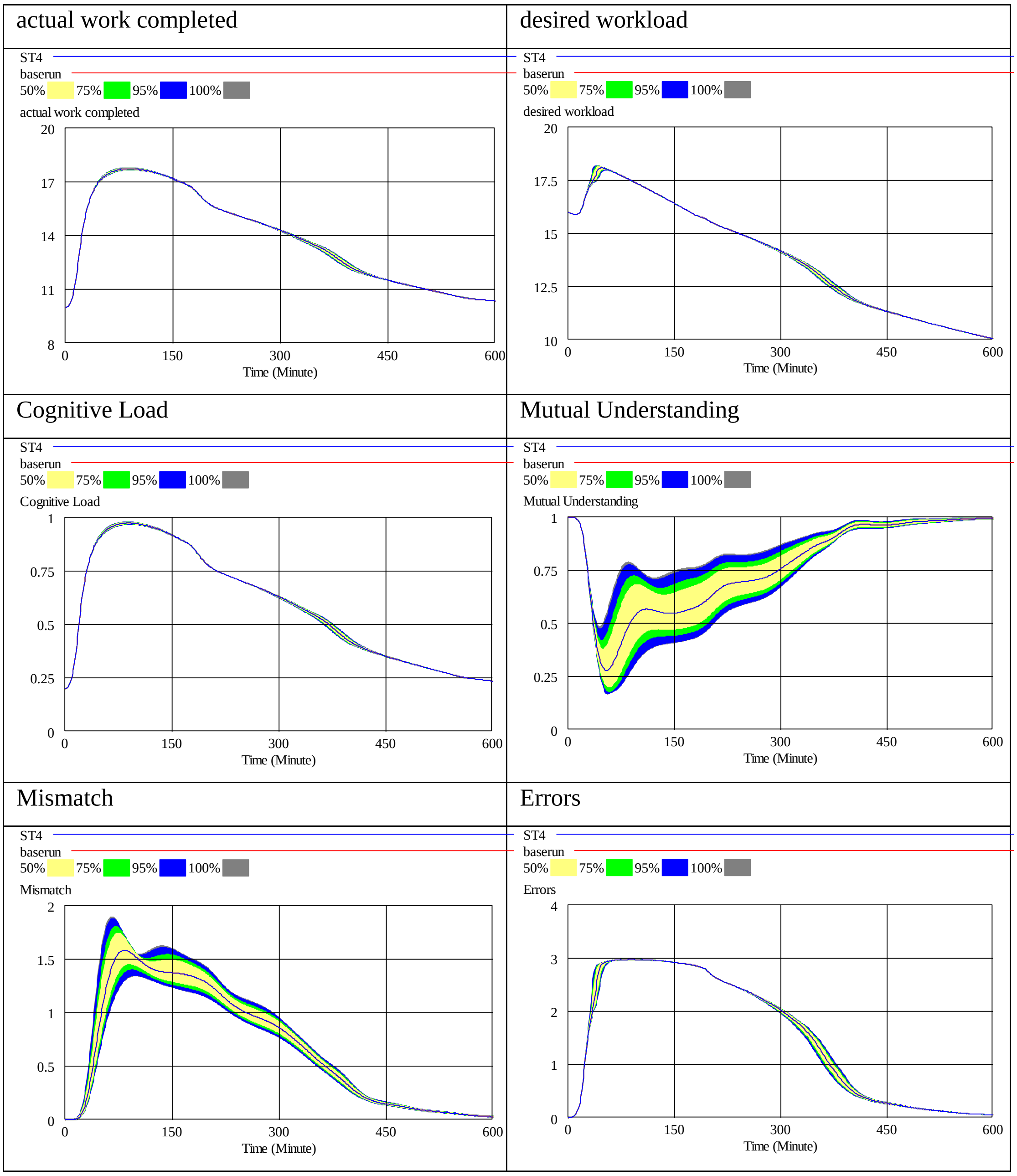


Figure 21 ST4- time to generate mismatch

Analysis:

The “time to generate mismatch” has little impact on the “actual work completed”, “desired workload”, and “Cognitive load”. However, “Mutual Understanding” and “Mismatch” are quite sensitive to “time to generate mismatch”. This is because “time to generate mismatch” directly affects “mismatch generated”. When the “time to generate mismatch” is long, the “mismatch generated” is small. “Mismatch” will be smaller than it otherwise would be. When “Mismatch” is small, less “mismatch corrected” occur, and the “Mutual Understanding” is smaller than it otherwise would be.

Numerical sensitivity is observed in “Mismatch” and “Mutual Understanding”. But no pattern sensitivity found. The numerical sensitivity is reasonable regarding the structure of the system.

Result: Pass

ST5- time to increase MU (mutual understanding)

Base run: Time to increase MU = 30

Sensitivity test: Range: 20-40; Distribution: Random Uniform; Runs: 200;

Simulation behavior:

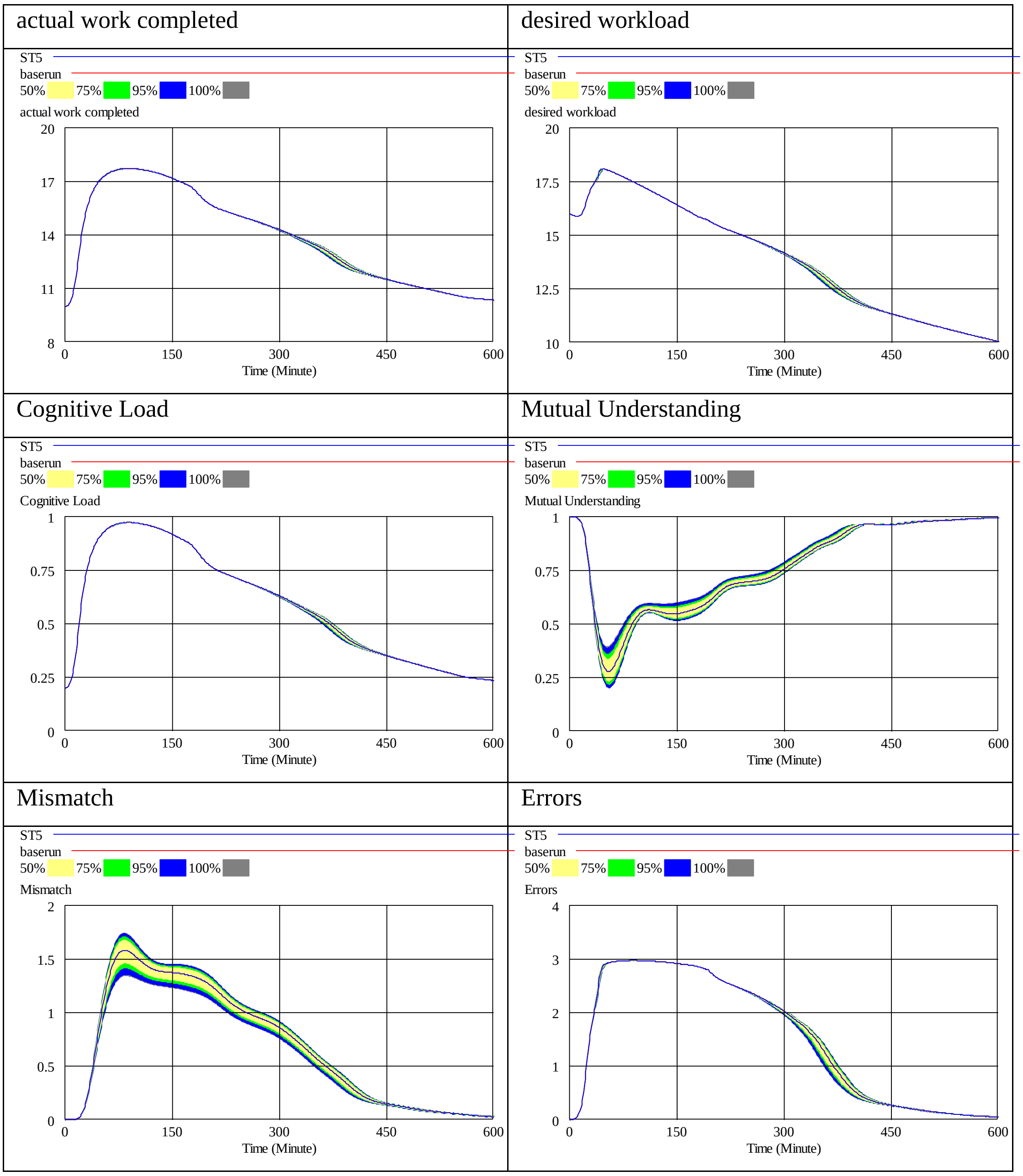


Figure 22 ST5- time to increase MU

Analysis:

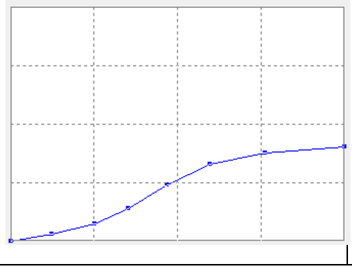
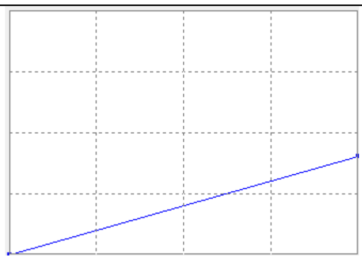
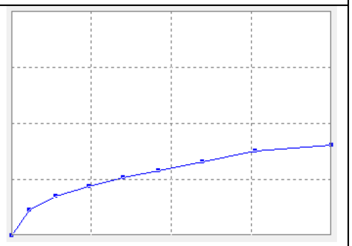
The “time to increase MU” has little impact on the “actual work completed”, “desired workload”, “Cognitive load”, and “Errors”. However, “Mutual Understanding” and “Mismatch” show a little sensitivity to “time to increase MU”. This is because “time to increase MU” directly affects “Mismatch”. When the “time to increase MU” is long, the “Mutual Understanding” increases slowly. “Mutual Understanding” will be smaller than it otherwise would be. As “Mutual Understanding” directly affects “Mismatch”, “Mismatch” shows similar sensitivity.

Numerical sensitivity is observed in “Mismatch” and “Mutual Understanding”. But no pattern sensitivity found. The numerical sensitivity is reasonable regarding the structure of the system.

Result: Pass

S6- Effect of local innovation on MU

This sensitivity test is to test the lookup function that lookup “Cumulative Work Rate Improvement” (generated by local innovation) to determine Mutual understanding. The lookup function in the base run is an S-shape curve, we test what will happen if the look up function is set up as linear and a convex curve.

<table>
<tr><td>Baserun</td><td>S61</td><td>S62</td></tr>
<tr><td></td><td></td><td></td></tr>
</table>

<u>Simulation behavior</u> (Baserun: blue line, No. 1; S61: red line, No. 2; S62: green line, No. 3)

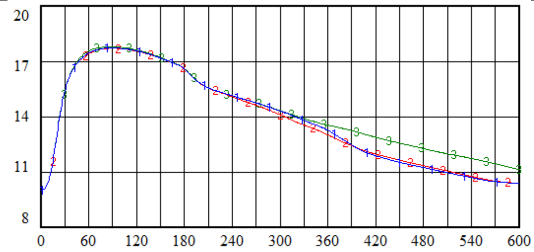
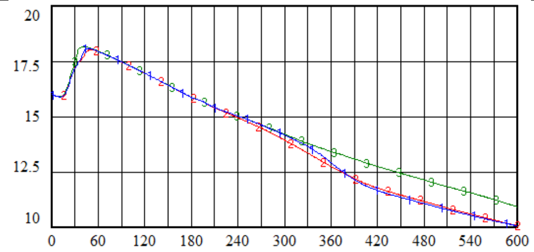
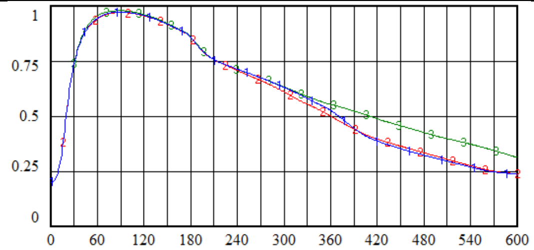
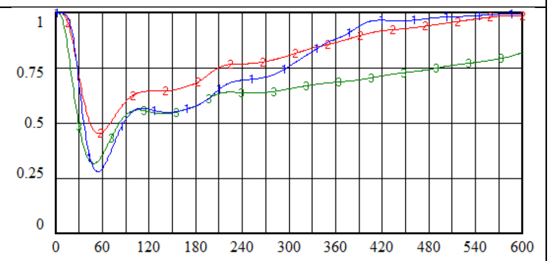
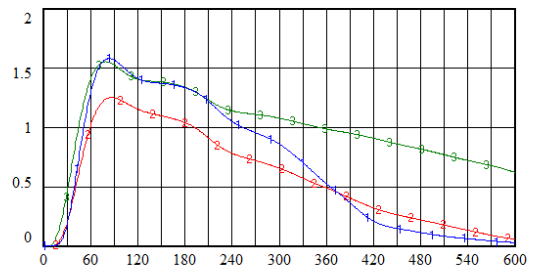
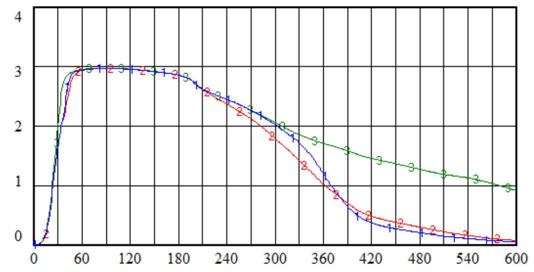
<table>
<tr><td>actual work completed</td><td>desired workload</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>Cognitive Load</td><td>Mutual Understanding</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>Mismatch</td><td>Errors</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

Figure 23 S6- Effect of local innovation on MU

<u>Analysis</u>:

Change in the “Effect of local innovation on MU” have little impact on “actual work completed”, “desired workload”, “Cognitive Load”. “Mutual Understanding”, “Mismatch” and “Errors” show some sensitivity in the later stage of the behavior.

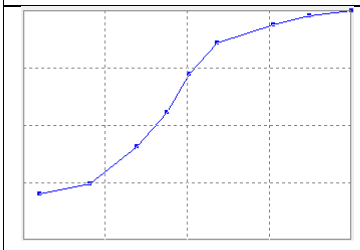
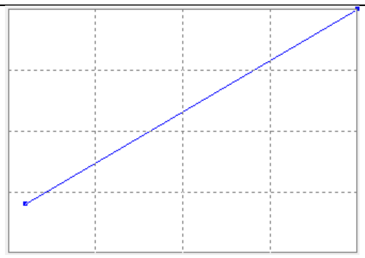
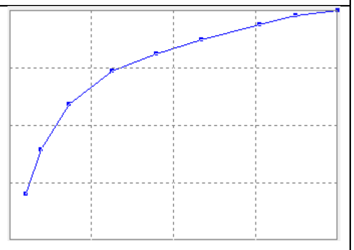
The sensitivity displayed in the model behavior is reasonable.

<u>Result</u>: Pass

S7- Indicated cognitive load

This sensitivity test is to test the lookup function that lookup “work load backlog” to determine cognitive load.

The lookup function in the base run is an S-shape curve, we test what will happen if the look up function is set up as linear and a convex curve.

Baserun	S71	S72
		

<u>Simulation behavior</u> (Baserun: blue line, No. 1; S71: red line, No. 2; S72: green line, No. 3)

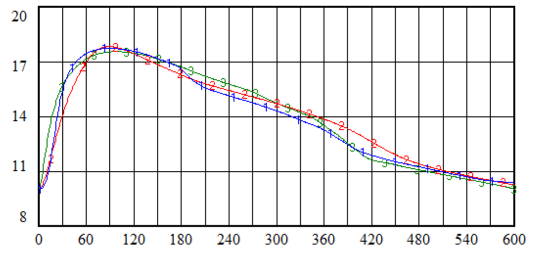
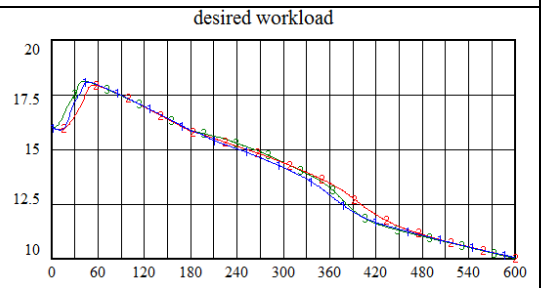
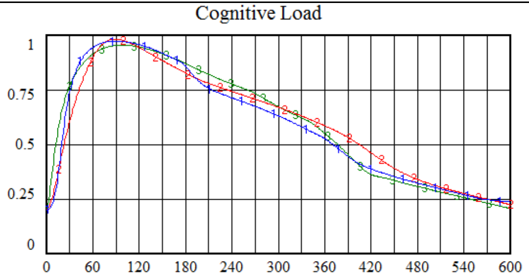
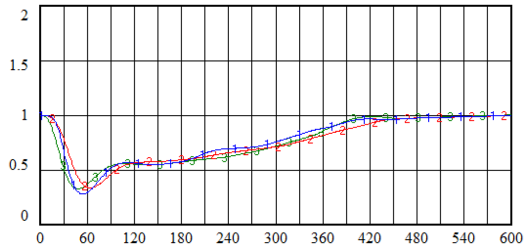
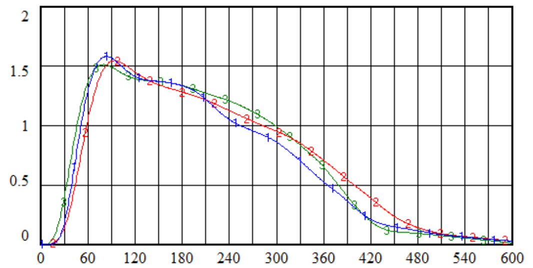
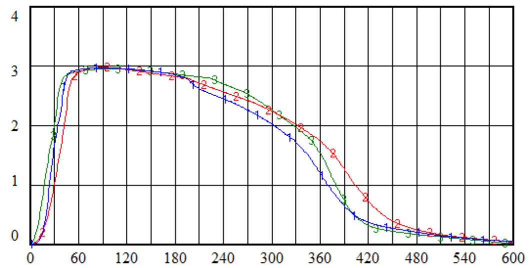
actual work completed	desired workload
	
Cognitive Load	**Mutual Understanding**
	
Mismatch	**Errors**
	

Figure 24 S7- Indicated cognitive load

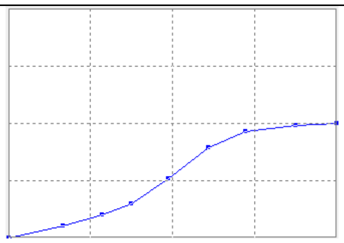
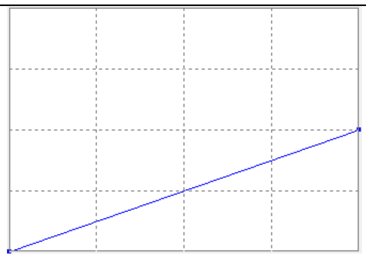
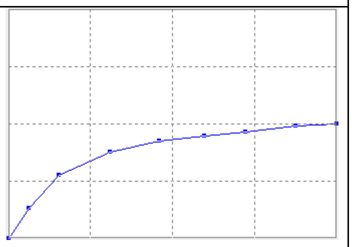
<u>Analysis</u>:

Change in the "Indicated cognitive load" have little impact on "actual work completed", "desired workload", "Cognitive Load", "Mutual Understanding", "Mismatch" and "Errors" show little sensitivity. The sensitivity displayed in the model behavior is reasonable.

<u>Result</u>: Pass

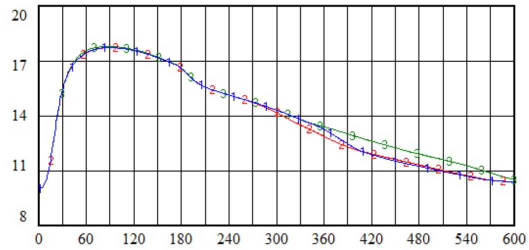
S8- Errors rate from mismatch

This sensitivity test is to test the lookup function that lookup “mismatch” to determine the errors rate from mismatch. The lookup function in the base run is an S-shape curve, we test what will happen if the look up function is set up as linear and a convex curve.

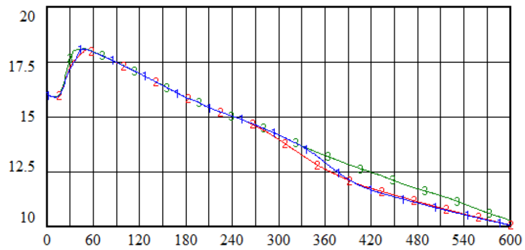
Baserun	S81	S82
		

Simulation behavior (Baserun: blue line, No. 1; S81: red line, No. 2; S82: green line, No. 3)

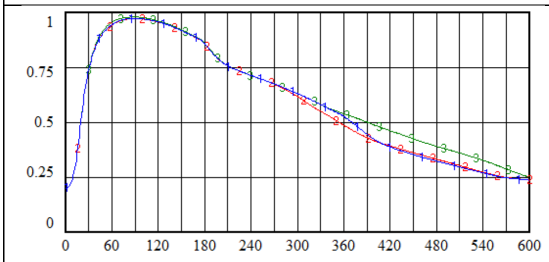
actual work completed



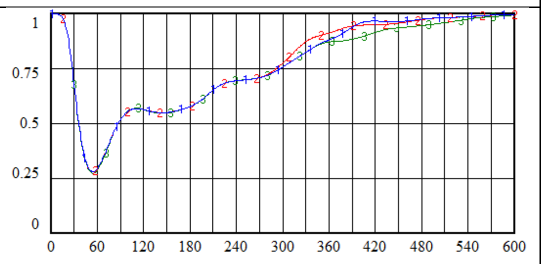
desired workload



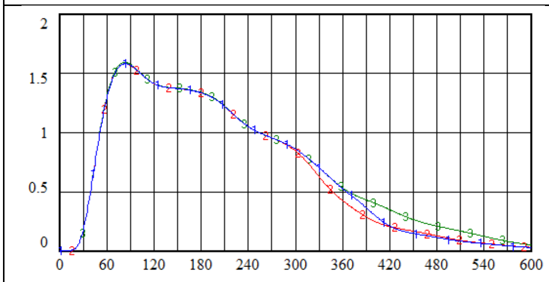
Cognitive Load



Mutual Understanding



Mismatch



Errors

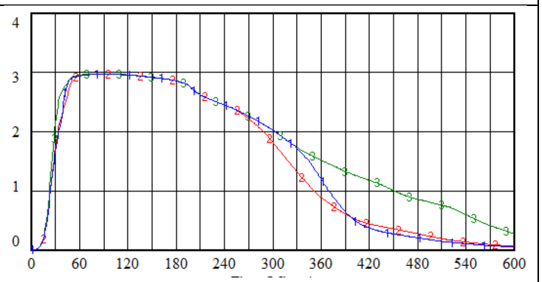


Figure 25

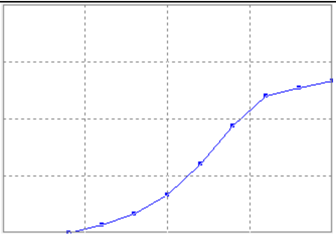
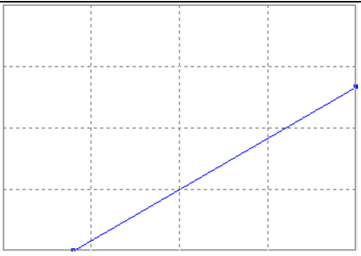
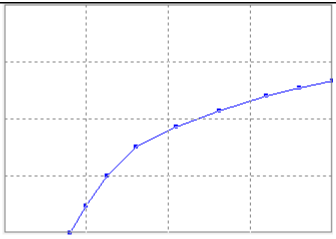
Analysis:

Change in the “Errors rate from mismatch” have little impact on “actual work completed”, “desired workload”, “Cognitive Load”, “Mutual Understanding”, “Mismatch” and “Errors” show little sensitivity. The sensitivity displayed in the model behavior is reasonable.

Result: Pass

S9- Errors rate from cognitive pressure

This sensitivity test is to test the lookup function that lookup “cognitive pressure” to determine the errors rate from cognitive pressure. The lookup function in the base run is an S-shape curve, we test what will happen if the look up function is set up as linear and a convex curve.

Baserun	S91	S92
		

Simulation behavior (Baserun: blue line, No. 1; S71: red line, No. 2; S72: green line, No. 3)

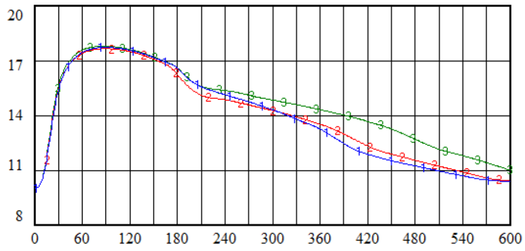
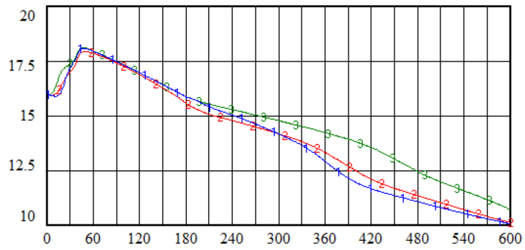
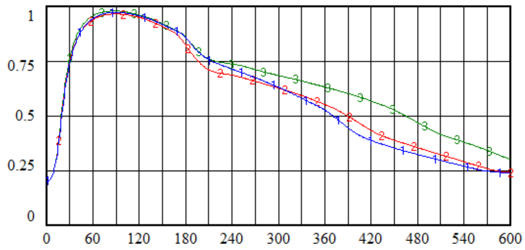
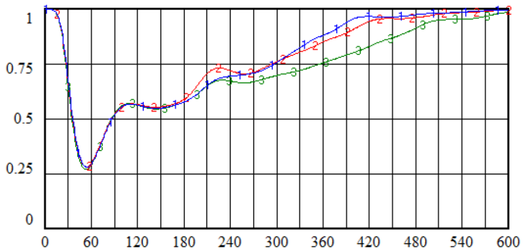
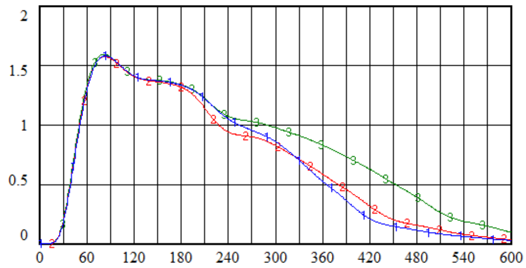
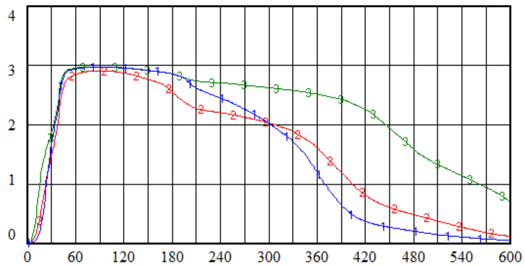
actual work completed	desired workload
	
Cognitive Load	Mutual Understanding
	
Mismatch	Errors
	

Figure 26 S9- Errors rate from cognitive pressure

Analysis:

Change in the “Errors rate from cognitive pressure” have little impact on “actual work completed”, “desired workload”, “Cognitive Load”, “Mutual Understanding”, “Mismatch” and “Errors” show little sensitivity. The sensitivity displayed in the model behavior is reasonable.

Result: Pass

CONCLUSION AND DISCUSSION

The agreement of the simulated behavior with the reference behavior and the fact that the simulation model passes a rich series of tests encourages us to suggest that the system dynamics model proposed in this paper embodies a rudimentary middle-range theory for the transition from disorganization to self-organization in emergencies. We suggest that this rudimentary theory is worth exploring as a starting point to gain more, and much needed understanding of the management of disorganization in emergencies. To establish a strong link to working hypothesis used by practitioners and a unified grand theory of management of emergencies more and better data is needed.

Forrester (1980) argued that models in social science must use all three kinds of existing data, viz., data stored mentally in people's heads (mental data), data stored descriptively in writing (written data), and data available numerically (numerical data). The numerical data are a tiny fraction of what is found in written form, which again is tiny compared to what people have in their heads. Excluding essential mental data would be tantamount to trying to manage a business while ignoring perceptions, beliefs, sentiments, and all other key data upon which decisions must be made.

The huge amount of mental data on emergency preparedness and response owned by practitioners is largely not available for scientists. Instead, the less abundant written data, and the even less abundant numerical data shape most of the current research on emergencies. By showing that even modest knowledge about the reference behavior of soft variables can facilitate theory building, this paper aspires to motivate practitioners to share more of their mental data and to inspire researchers to include direct questions to the practitioners about data series, that is, reference behavior for even the softest of the variables in forensic studies of emergencies. In particular, the power of Delphi techniques should be used to obtain such reference behavior.

ACKNOWLEDGMENTS

We are very grateful for valuable comments and suggestions from Starr Roxanne Hiltz, Murray Turoff, Geir Bøe, John Einar Johansen and Shengxin Chen.

REFERENCES

Boin, A. and P. t. Hart (2008). Public Leadership in Times of Crisis: Mission Impossible? Crisis management. A. Boin. Los Angeles, Sage**:** 1-15.

Byrne, D. S. (1998). Complexity theory and the social sciences : An introduction. London, Routledge.

Dynes, R. R. and E. L. Quarantelli (1976). Organizational Communications and decision Making in Crises. Newark, DE, USA, Report Series 17, University of Delaware, Disaster Research Center.

Forrester, J. W. (1980). "Information Sources for Modeling the National Economy." Journal of the American Statistical Association **75**(371): 555-566.

Gilpin, D. R. and P. J. Murphy (2008). Crisis management in a complex world. Oxford, Oxford University Press.

Gonzalez, J. J., G. Bøe, et al. (2013). A System Dynamics Model of the 2005 Hatlestad Slide Emergency Management 10th International Conference on Information Systems for Crisis Response and Management. Baden-Baden, Germany.

Hutchins, E. (1995). Cognition in the Wild. Cambridge, Mass., MIT Press.

Jasny, B. R., G. Chin, et al. (2011). "Again, and Again, and Again …." Science **334**(6060): 1225.

Lane, D. C. (2008). "Formal theory building for the avalanche game: Explaining counter-intuitive behaviour of a complex system using geometrical and human behavioural/physiological effects." Systems Research and Behavioral Science **25**: 521-542.

Lango, P. (2010). Samordning i krise eller krise i samordning? Bergen, University of Bergen. **Master thesis**.

Lango, P. (2011). Hatlestad-raset. Organisering, samfunnssikkerhet og krisehåndtering. A. L. Fimreite, P. Lango, P. Lægreid and L. H. Rykkja. Oslo, Universitetsforlaget**:** 216-234.

Maani, K. E. and R. Y. Cavana (2007). Systems thinking, system dynamics : Managing change and complexity. North Shore, N.Z., Pearson/Prentice Hall.

Macintosh, R. and D. Maclean (1999). "Conditioned emergence: A dissipative structures approach to transformation." Strategic Management Journal **20**(4): 297-316.

Merton, R. K. (1968). Social Theory and Social Structure. New York, Free Press.

Meteorological Institute Oslo (2005). Varsling av stormer og ekstremt vær, 18/2005. Varsling av stormer og ekstremt vær. Oslo, Norway.

Multiconsult AS (2005). Notat G-02. Rasrisiko etter rekordnedbør 14. september. Hatlestad. Stabilitetsforhold. Årsaker til skredet. Bergen

Palttala, P., C. Boano, et al. (2012). "Communication Gaps in Disaster Management: Perceptions by Experts from Governmental and Non-Governmental Organizations." Journal of Contingencies and Crisis Management **20**(1): 2-12.

Richardson, G. P. and A. L. Pugh (1981). Introduction to system dynamics modeling. Portland, Or., Productivity Press.

Schwaninger, M. and S. Grösser (2008). "System Dynamics as model-based theory building." Systems Research and Behavioral Science **25**: 447-465.

Simpson, N. C. and P. G. Hancock (2009). "Fifty years of operational research and emergency response." Journal of the Operational Research Society **60**(Supplement 1): 126-139.

Sterman, J. D. (2000). Business dynamics : Systems thinking and modeling for a complex world. Boston, Irwin McGraw-Hill.

Tu, Y.-m., W.-y. Wang, et al. (2009). "The essence of transformation in a self-organizing team." System Dynamics Review **25**(2): 135-159.

Turoff, M., J. J. Gonzalez, et al. (2012). Modeling organizational behavior in emergencies. NOKOBIT, Bodø, Norway, Akademika Forlag.

Wikipedia (2007). "Hatlestad Slide, http://en.wikipedia.org/wiki/Hatlestad_Slide." Retrieved March 16, 2014, from http://en.wikipedia.org/wiki/Hatlestad_Slide.

Yin, R. K. (2009). Case study research : design and methods. Thousand Oaks, Calif., Sage.