Biased Decision Making in Realistic Extra-Procedural Nuclear Control Room Scenarios

Abstract: In normal operations and emergency situations, operators of nuclear control rooms rely on procedures to guide their decision making. However, in emergency situations, where several interacting problems can cause unpredictable adverse effects, these procedures may be insufficient in guiding operators to safe shutdown of the power plant. However, little is known about the decision making strategies that operators employ in these extra-procedural situations. To address this, a realistic simulation study was conducted with five crews of active, licensed nuclear operators to see the behavioural patterns that occur when procedures are not sufficient. This paper is a re-analysis of a dataset collected for a different study, aiming to investigate how the design and existence of procedures influence, and possibly bias, decision-making strategies. We found evidence that operators were affected by confirmation bias, and that, in some cases, the mismatch between their home power plant and the simulated power plant made them commit errors due to misapplied expertise. We further found that this effect was amplified by the existence and design of the procedures used. Based on these findings, designers we suggest that designers may improve safety by creating procedures that bear the risks of these biases in mind, or by specifically aiming to debias the users. Avenues for debiasing through design are discussed.

1. Introduction

Studies of operators in nuclear control rooms, airplane cockpits and medical decision making have led to a greater understanding on decision making in high-stakes complex environments over the last several decades. To deal with the complex requirements of these environments, researchers in decision making and design have worked hand in hand to improve the interfaces, environments and tools of the specialists that operate these fields to improve performance (e.g. [1]–[5]). An important early development was the move towards the use of written procedures and checklists. Procedures and checklists are written documents, usually physical paper copies of varying length, that specify conditions for their use, followed by a list of diagnosis and action steps. However, previous studies have found that for the critical situations in nuclear control room, the majority of real life operating events included non-typical conditions [6], [7]. In such situations, where the predicted situation in the procedure does not match the observed situation, procedures may become inefficient or lack proper guidance [8].

In the design field, an increasing amount of studies have sought uncover patterns in human behaviour in order to guide designers in their efforts, on topics ranging from basic perceptual functioning, to aesthetic product preferences and other critical factors for consumer decisions (e.g.[9]–[17]. Through evaluations of real and/or stylized products and product attributes, these studies have started to shape our understanding of how humans, in general, and in relation to specific groups, perceive and interact with various design characteristics. Furthermore, studies of designers have found biases in their decision making such as design fixation([18], for a recent review see [19], the preference effect [20], strategies for how these effects can be mitigated, and how these strategies interact with expertise [21]. Outside of the design field, developments over the last several decades have led to an increased understanding of decision making strategies and biases in general [22], [23]. However, little is known about how these decision making insights apply in the practical situation of a control room emergency, and whether the decision making biases are reduced or amplified by the existence of the designed objects such as procedures and checklists – particularly in the non-typical situations that characterize real life emergencies.

To address this, the present paper re-analyses data collected for a project involving two realistic pressurized water reactor scenarios conducted at the Halden Human Machine Laboratory (HAMMLAB) simulator in 2014. The scenarios were designed such that multiple complications would lead to situations where crews had to perform autonomous extra-procedural actions to achieve optimal performance. The results of the original study were documented by Massaiu & Holmgren [24]. They investigated how operators perceive discrepancies between their own plans and the procedure, how crews compromised between needing to act fast and to follow procedures, and how the crew size and composition affected diagnosis and decision-making. They found that crews, with some exceptions, prioritized strict adherence to

Please cite as:
Andersen, E., Kozin, I., & Maier, A. (2018). Biased Decision Making in Realistic Extra-Procedural Nuclear Control Room Scenarios. In *Design Computing and Cognition DCC'18* (pp. 601-620). Springer. https://doi.org/10.1007/978-3-030-05363-5_30

procedures and that crew size and composition did not influence performance. The scenarios were described in detail by Massaiu & Holmgren [25] to allow for future re-analysis, such as this paper. Adding to this former work, the present paper aims to show biases and heuristics that may have caused divergences in behaviour amongst the crews. The purpose of this study then is to create an exploratory platform to show how biases may influence expert decision making in critical situations, as well as start shaping our knowledge of how these biases may arise from the designed objects, that these operators interact with during their work.

2. Background

In this paper, we focus on two biases that have been related to expert decision making: The first is the bias that occurs when expertise is transferred to a similar, but different situation, thus causing misapplications of one's expertise. The second is confirmation bias, which is the tendency to overly prioritize and seek for information that benefits existing views. Both biases have been shown to impact decision making of experts many diverse fields, such as medicine, engineering and law [26]

In this section, these biases are described, and we outline which behavioural patterns should be observed if the nuclear control room operators are affected by them.

2.1. Expertise

2.1.1. What is an Expert?

In this paper, we use as the basis for our definition of expertise the one given by Simon [27]: "The situation has provided a cue: This cue has given the expert access to information stored in memory, and the information provides the answer. Intuition is nothing more and nothing less than recognition". From this viewpoint, an expert is one who has been exposed to a high variety of situations and has learned the correct response, which allows him to swiftly recall and apply it in future situations. Furthermore, we extend our definition of expertise based on the arguments by Kahneman & Klein [22]. Drawing on a study by Shanteu [28], they argue that expertise will only form if a) the context of training provides valid cues for learning, meaning cues that reflect real patterns in the context, and b) if the context of training is sufficiently regular to allow for learning of patterns. Without these aspects, they argue, it is not possible to learn whether your behaviour is resulting in good or bad results, and thus expertise cannot be achieved. The nuclear power plant is a vastly complex system, with a myriad of technical details

Please cite as:
Andersen, E., Kozin, I., & Maier, A. (2018). Biased Decision Making in Realistic Extra-Procedural Nuclear Control Room Scenarios. In *Design Computing and Cognition DCC'18* (pp. 601-620). Springer. https://doi.org/10.1007/978-3-030-05363-5_30

that needs to be acquired over several years an operator is certified. However, from the criteria of Kahneman & Klein [22], the nuclear power plant control room is a valid context for acquiring expertise, as the relation between inputs and outputs of actions is both consistent and readily available of observation for the operators. Given the extensive training required for certification, as well as the substantial experience of all the participants in the study, a high level of expertise should thus be expected.

2.1.2. Misapplied Expertise

Given this expertise, it is expected that the operators employ highly refined heuristics (short-cuts for decision making) that allow them to make (near-) optimal decisions for the context they have been trained in, with lower effort [23], [29], [30]. However, these strategies can decrease performance if applied to other contexts, where they are not adequate. For the present simulation study, this may be the case. First, the crews were trained at a power plant in a different country than the simulated power plant. Second, not all parts of the simulation perfectly matched what would be observed in the reference power plant. The operators could perform suboptimally due to lack of plant specific knowledge or due to expectations of and/or reliance on signals that do not come due to plant differences. Furthermore, due to these differences, the operators need to adjust their behaviour to reflect a lower level of expertise than what they have for their home plant, taking the more explorative mind-set of a novice. However, previous research has shown that experts, when put in a similar, but not identical, situation to what they have expertise for tend to act as if their expertise applies to the novel situation as well [22]. This is caused by a false belief that there is a perfect transfer of skill between the two situations. While the operators receive training in operating the simulation power plant prior to the simulation scenarios, there may nevertheless be deviations between the two power plants that will cause operators to use heuristics that are inappropriate for the specific context.

If the nuclear power plant operators are affected by the bias of misapplied expertise, we expect that they will insufficiently double-check their decisions (as they would not need to do this if they were highly trained) and to make deviations that turn out wrong due to lacking plant specific knowledge.

2.2. Confirmation Bias

Confirmation bias is the non-conscious tendency to seek for- and give higher value to information that confirms our existing views, and, conversely, to ignore and deprioritize information that goes against our existing views. Confirmation bias is thus an overarching term that covers tendency to strongly persist in existing beliefs, as a result of biased evaluation of information and in search for information [26].

2.2.1. Belief persistence

The first aspect of confirmation bias is belief persistence, which is the term for a collection of tendencies that cause early beliefs to be very resistive to change: First, the tendency to persist in early hypotheses for no reason than them being the first adopted hypotheses [31]. Second, the tendency to be more likely to question information that contradicts their existing belief, while being less likely to question information that confirms their pre-existing belief [32], [33]. Third, the tendency to be likely to explain away events as random etc. if they conflict with their existing beliefs, thus discrediting the events rather revisiting the belief [34].

If operators are susceptible to belief persistence, we expect that operators will persist in their early hypotheses if they are do not contradict the operating procedures (regardless of whether or not the procedures are correct for optimal decision making at the time).

2.2.2. Biased Search for Information

The second aspect of confirmation bias is the tendency to only seek information that confirms one's existing view, or to only seek for information that would only exist if the existing view was correct. Conversely, it is the tendency to avoid information that would disconfirm one's view and/or not to seek for information that would exist if an alternate view that was correct [35]. This tendency thus allows one to never disconfirm one's view through never exposing oneself to situations that threaten the viewpoint. Furthermore, given that one only samples information that supports the view, confidence in the view increases [36].

Similarly, we expect that operators will perform confirmatory search by looking at power plant locations that will show problems only if their hypothesis was true, and will tend not to search for disconfirming information through e.g. alternate sources such as field operators.

3. Case Study

The data that form basis for this paper are the decisions of nuclear control room operators in two realistic simulation scenarios, conducted in the HAMMLAB simulator in 2014. Two scenarios of realistic emergency situations in a Pressurized Water Reactor were run by five crews of 3-5 crew members. The size of the crews and the exact scenario durations are shown in Table 1. In nuclear operations, operators rely on Emergency Operating

Please cite as:
Andersen, E., Kozin, I., & Maier, A. (2018). Biased Decision Making in Realistic Extra-Procedural Nuclear Control Room Scenarios. In *Design Computing and Cognition DCC'18* (pp. 601-620). Springer. https://doi.org/10.1007/978-3-030-05363-5_30

Procedures to solve emergencies. The operators have knowledge of a vast array of 'entering conditions' for various procedures, and will 'enter' a given procedure in response to these conditions. Once entered, the procedures will guide operators through identifying and alleviating problematic symptoms. Operators are at no point required to know the cause of the observed problems, only to follow the procedures to alleviate the results of these symptoms.

Table 1: Crew size and duration of scenarios. Note that shorter duration does not indicate better performance in all cases.

Crew	1	2	3	4	5
Size	5	5	3	4	3
SC#1 Duration	02:06:21	01:54:47	02:50:00	01:52:51*	02:10:37
SC#2 Duration	01:26:21	01:14:47	02:10:00	01:12:51*	01:30:37

*Scenario was stopped before the crew had completed the final goal

The two scenarios are unique with respect to the cause of the problem. However, in both scenarios, emergency operating procedures are entered in response to the reactor 'tripping' (this term refers to neutron absorbing control rods being inserted into the core, thus stopping chain reaction). While tripping the reactor stops further power from being produced, the power plant is not safe until problems such as leaks causing spread of radiation are solved, and the plant is cooled and depressurized. Until safe shutdown is achieved, adverse effects such as release of radioactive material to the atmosphere, or, in the worst case, core meltdown, are still possible.

Both scenarios were designed such that following operating procedures was not sufficient for safe and effective shutdown. Operators were thus required to perform autonomous actions to avoid adverse effects. The scenarios are described in detail below. Overall, the complex problems of the scenarios caused several problems for all crews in both scenarios, albeit to varying degrees for the various crews, as will be elaborated below.

3.1. Scenario #1

The first scenario involved multiple leaks on the piping system that connects the core with the plant's three steam generators. The crew is given the cover story that construction is ongoing nearby, which, shortly after start, is cited as the cause for a blast that gives vibrations to the plant, including the control room. In the simulation, this blast results in immediate release of radioactive material in all steam generators, followed shortly (12 mins after

Please cite as:
Andersen, E., Kozin, I., & Maier, A. (2018). Biased Decision Making in Realistic Extra-Procedural Nuclear Control Room Scenarios. In *Design Computing and Cognition DCC'18* (pp. 601-620). Springer. https://doi.org/10.1007/978-3-030-05363-5_30

start) by a small leak in a tube connected to steam generator #2, and subsequently (20 mins after start) a rupture in a tube in steam generator #3. The leak in steam generator#3 will increase in size two times, first at 25mins and then at 40mins after start, with the latter being equivalent of a complete tube rupture. If the crew has not manually tripped the reactor at 40mins, the automatic tripping system will do so shortly after the complete tube rupture.

The challenge for the crew is ensure cooldown while avoiding using the two damaged steam generators' relief valves, as this would result in release of radioactive material to the atmosphere. To do so, the two damaged steam generators should be isolated. This task is complicated by the fact that it is not clear from following the procedures and the information displayed whether there whether steam generator 2 is causing problems, as it is obscured by the effects of the rupture in steam generator 3. Operators must thus actively look for additional information to successfully handle the task.

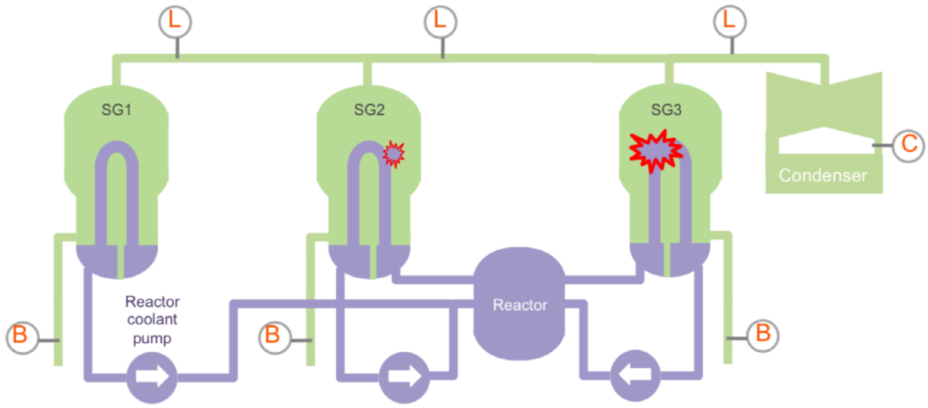

Figure 1: A simplified diagram of scenario 1. First, a leak occurs in Steam Generator 2. Second, a leak occurs in Steam Generator 3, which develops to a full tube rupture over 40 minutes. Diagram courtesy of Massaiu & Holmgren [24].

3.2. Scenario #2

The second scenario involves an irreversible loss of coolant following leaks to the Reactor Coolant System, which results in water spilling on the floor of the auxiliary building. The scenario begins with a distracting task in the form of a pump trip. This will occupy the operators at the start of the scenario. The first major complication happens when two valves starts leaking in the Residual Heat Removal System (one at start, the other after 8 minutes). At around 11 mins from start, a pipe in the Residual Heat Removal system of the auxiliary building will break, resulting in reactor coolant fluid spilling on the floor. Finally, a smaller leak will occur in the Reactor Coolant

Please cite as:
Andersen, E., Kozin, I., & Maier, A. (2018). Biased Decision Making in Realistic Extra-Procedural Nuclear Control Room Scenarios. In *Design Computing and Cognition DCC'18* (pp. 601-620). Springer. https://doi.org/10.1007/978-3-030-05363-5_30

Pump thermal barrier, which will complicate the detection of the primary leak. The loss of pressure will cause an automatic trip of the reactor if it is not initiated manually.

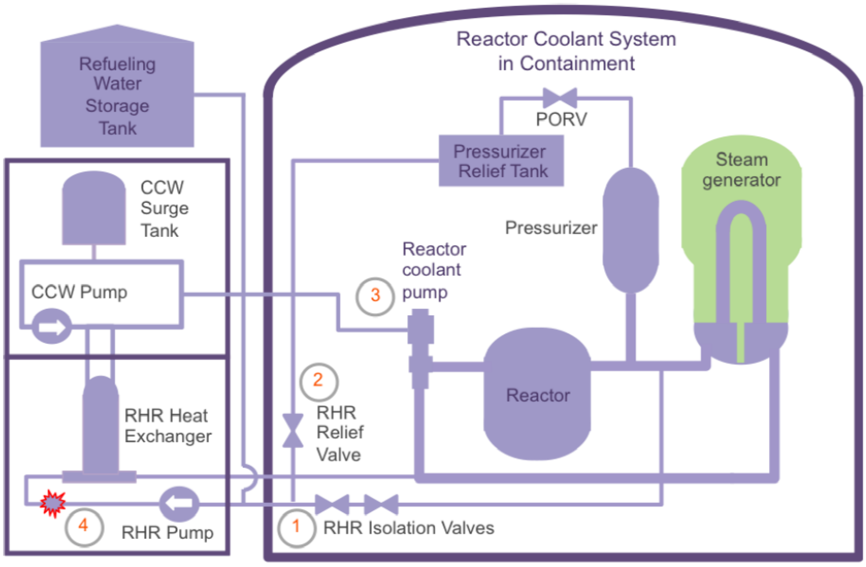

Figure 2: Simplified Diagram of Scenario 2. First, the Residual Heat Removal (RHR) isolation valves begin to leak. Second, the RHR relief valve opens and direct primary coolant to the Pressurizer Relief Tank (PRT). Third, a tube leaks in the Reactor Coolant Pump thermal barrier exchanger. Fourth, a pipe breaks in the RHR, releasing large amounts of water on the auxiliary building floor. Diagram courtesy of Massaiu & Holmgren [24].

The challenge for the crew is to ensure safe and effective cooldown while reducing the effects of the leaks. This task is complicated by the fact that the normal procedure for this type of event is to identify the leaks and to isolate them, whereas the leaks in the present scenario are not isolatable. Furthermore, operators do not have a clear indication that the leaks are not isolatable, and, it is considered optimal performance to look for the leaks, despite the procedures not requiring this.

4. Methods

In the following, we describe the dataset and analysis method, including the collection site, the Operator Performance Assessment System (OPAS, REF), which served as the measure of performance.

Please cite as:
Andersen, E., Kozin, I., & Maier, A. (2018). Biased Decision Making in Realistic Extra-Procedural Nuclear Control Room Scenarios. In *Design Computing and Cognition DCC'18* (pp. 601-620). Springer. https://doi.org/10.1007/978-3-030-05363-5_30

4.1. Dataset

Five crews of certified operators from three nuclear power plants were recruited. The crews varied in size: Two crews had three members, two crews had four members and one crew had five members. All crews participated in both scenarios.

The study took place at the Human Machine Laboratory (HAMMLAB) at the Institute for Energy Technology, Halden, Norway, in a realistic simulation set-up that mimics a Swedish Pressurized Water Reactor (PWR). Audio and video materials were recorded during the scenarios, which serve as the raw dataset for this analysis. The audio material included all conversation between operators, sounds played in the environment, conversations between the operators and the experimenters (who, at various times, roleplayed as field operators) and conversations between experimenters. The video material consisted of four streams of serially played still-shots of the operators (one on each operator and an overview camera) and recordings of the displays used by the operators (including mouse movements on these).

The raw dataset was processed by a nuclear power plant expert, who has many years of experience as an operator and as a trainer of operators, and thus has the required skills to evaluate performance. These evaluations and additional comments were noted down in a detail in a MS Excel Spreadsheet, along with the timestamp and a brief description of the various events that the process expert had deemed significant. The contents of the comments included, but were not limited to: examples of good behaviours, errors and deviations from protocols.

To measure performance, each team was evaluated by the process expert using the OPAS system [37], which gives a score of one to five on task critical operations, and with five being perfect performance. For the purpose of this study, the OPAS scores for each scenario were averaged as an indicator of overall performance.

As this study re-analyses previous data, the dataset for this paper includes the aforementioned raw data and processed MS Excel Spreadsheet, as well as a detailed report with further processed descriptions of the events [25].

4.2. Analysis Procedure

To perform the analysis for this report, the first author first read the report by Massaiu and Holmgren [25], which gives a detailed description of the scenarios for each crew. This was to better familiarize him with the terminology and to look for possible connections to theory. Second, he analyzed the detailed MS Excel spreadsheets, marking events that were of particular interest for further processing. Third and final, he re-inspected the video and audio material to find examples of specific dialogue exchanges, and to listen

Please cite as:
Andersen, E., Kozin, I., & Maier, A. (2018). Biased Decision Making in Realistic Extra-Procedural Nuclear Control Room Scenarios. In *Design Computing and Cognition DCC'18* (pp. 601-620). Springer. https://doi.org/10.1007/978-3-030-05363-5_30

for cues, such as tonality, formality of word usage, volume in the speech to get a better understanding of the marked events. The procedure for re-inspecting an event was to start the recording approximately one minute before the timestamp of the event, and to end approximately one minute after the event had ended, in order to ensure that the context was properly understood. The purpose of the re-inspection was to ensure that the writing of the process expert (which did not contain comments about tonality etc.) was not misunderstood, which was especially important when the process expert had written ad verbatim quotes from the operator dialogue.

Following the re-inspection of the data, the first author marked down all examples of good behaviour (as commented by the process expert) in the aforementioned MS Excel spreadsheet with a green mark, and all examples of errors and deviations with a red mark. Following this, these behaviour examples were coded into categories, and the frequency of each category was counted to give an overview of the performances.

Based on these steps, behavioural patterns of the highest, medium, and lowest scoring crews (as defined by the OPAS score) were compared.

5. Results

In this section we show the behavioural findings from the simulation scenarios. The results are divided by scenario, and for each scenario we describe the behavioural patterns of the top-, medium- and lowest-scoring crew(s). The differences between these behavioural patterns, and the extent to which they are caused by use of procedures and decision making biases will be discussed in section 6.

5.1. Scenario 1

In scenario 1, the major point of divergence in team performance lied in identifying that both steam generator #2 and #3 suffered from structural damage. Two of five crews detected that there were problems with both steam generators, whereas the remaining three crews proceeded as if only one steam generator was leaking at any given time. However, all teams were challenged in this detection, as it adhered to procedure and was plausible that any effects observed from steam generator #2 could have been caused by ‘shine’ (which is the term for radioactive measures spilling over from a larger nearby rupture). The OPAS scores for scenario 1 are found in table 2. Problems were observed for all teams, and radioactive material was released to the atmosphere by all crews (crew 5 was the only crew who did so intentionally). As was previously reported by Massaiu & Holmgren [24], crew

Please cite as:
Andersen, E., Kozin, I., & Maier, A. (2018). Biased Decision Making in Realistic Extra-Procedural Nuclear Control Room Scenarios. In *Design Computing and Cognition DCC’18* (pp. 601-620). Springer. https://doi.org/10.1007/978-3-030-05363-5_30

size did not predict performance and the teams with a Shift Technical Advisor did not outperform teams without a person performing this role.

In the following, we characterize the differences between the top, medium and lowest scoring crews, by describing their behavioural patterns. Differences in these behavioural patterns will be discussed in concordance with our hypotheses in section 6.

Table 2, crew size, duration of scenario and performance scores (OPAS measures) for scenario 1. Note that duration does not necessarily equate better performance.

Scenario 1 OPAS					
Crew	#1	#2	#3	#4	#5
Size	5	5	3	4	3
Alarm Handling	5	4	4	3	5
Identification and Isolation	3	4	3	3	4,5
Cooldown	3	5	4	4	5
Depressurization	4	1	4	3	4,5
Stop Safety Injection	n/a	5	5	2	5
Pressure Balance	n/a	3	2,5	1	4
Average	3.8	3.7	3.8	2.7	4.7
Duration	02:06:21	01:54:47	02:50:00	01:52:51*	02:10:37

*Scenario was stopped before the crew had completed the final goal

5.1.1. Factors Causing High Performance

The highest scoring crew was crew 5, with an average OPAS score of 4.7. Four aspects were observed that may have caused this performance: First, the process expert observed few technical errors and/or instances of suboptimal execution. Second, the crew was very active in looking for alternate sources of information and in testing multiple hypotheses. For example, crew 5 was the only crew to autonomously ask for information about Steam Generator #2 integrity, follow up to ensure that they received the information and perform actions to isolating Steam Generator #2. The focus on looking for alternate sources, and investigating alternate hypotheses was also visible in the language used in strategy meetings, where crew members would use utterances such as *"ruptured steam generator or generators" "we can't tell WHICH steam generator"* to describe the problem. Third, the Unit Supervisor did not read aloud notes in procedures and was, along with the rest of the crew, very proactive in planning future actions. Specifically, the crew frequently called for status updates, wherein ideas were discussed and shared and plans were laid for future actions. Fourth, the Unit Supervisor in crew 5 would read the entire procedure aloud before any actions were taken,

Please cite as:
Andersen, E., Kozin, I., & Maier, A. (2018). Biased Decision Making in Realistic Extra-Procedural Nuclear Control Room Scenarios. In *Design Computing and Cognition DCC'18* (pp. 601-620). Springer. https://doi.org/10.1007/978-3-030-05363-5_30

whereas in other crews, the Unit Supervisor would read each procedural step one by one as they were completed.

5.1.2. Factors Causing Medium Performance

The medium-scoring crews were crew 1, 2 and 3, with average OPAS scores of 3.8, 3.7 and 3.8. Of these, one crew, crew 3, identified that both Steam Generator #2 and #3 had suffered structural damage. However, all three crews considered, to varying degrees, whether Steam Generator #2 had leaked as well. The different strategies for investigating these considerations were as follows: Crew 3 called a 'Field Operator' (roleplayed by the control room), but without opening sampling valves (thus not allowing for sampling), and did not call the field operator back after the sampling valves had been opened. After some deliberation in the control room, the 'Field Operator' decided to call the crew and share the information that both Steam Generator#2 and Steam Generator#3 showed radiation, after which the crew promptly performed actions towards isolating Steam Generator#2 as well. However, the crew used Steam Generator#2 for cooldown. Crew 1 and 2 discussed the possibility of a leak in Steam Generator#2, opened sampling valves and sent a 'Field Operator' to collect samplings. However, both crews did not follow up with the Field Operator and abandoned the hypothesis that Steam Generator#2 could be leaking as well. Compared to crew 5, the three middle scoring crews used singular terms about the steam generator problem early on, such as *"the ruptured steam generator" (crew 1)*. Notably, some medium scoring crews originally believed the measures in Steam Generator#3 were due to shine (when it had first leaked) and then changed their hypothesis to it being the Steam Generator#2 measures that were caused by shine. These crews thus effectively did not change the hypothesis that it was only one steam generator that was faulty – they simply changed their mind about which one it was. Finally, the medium scoring crews tended to search for information that backed up (rather than falsified) their views, as is reflected in language such as *"request chemistry back that up with local sample"(crew1).*

5.1.3. Factors Causing Low Performance

The lowest scoring crew was crew #4, with an average score of 2.7. Two aspects were observed that may have caused this performance: First, the process expert observed technical errors in executing the procedures. The effects of these errors caused ripple effects that made the scenario more and more complex, eventually resulting in the scenario being ended before depressurization was achieved. Second, we observed examples of inefficient communication between crew members. For example, the crew's Reactor

Please cite as:
Andersen, E., Kozin, I., & Maier, A. (2018). Biased Decision Making in Realistic Extra-Procedural Nuclear Control Room Scenarios. In *Design Computing and Cognition DCC'18* (pp. 601-620). Springer. https://doi.org/10.1007/978-3-030-05363-5_30

Operator and Shift Technical Advisor suggested several times to perform steps to test whether Steam Generator #2 was also leaking. However, the Unit Supervisor was of a different belief, and thus did not translate the recommendations into actions towards isolation. As a result, the crew quickly abandoned the possibility that two Steam Generators were damaged and instead focused on Steam Generator #3.

5.2. Scenario 2

In scenario 2, the major point of divergence was the degree to which teams chose to invest resources into identifying the leak locations by going outside of the procedures. These differences are detailed below. Overall, the impact of choosing to use resources to identify and isolate the leaks varied: Teams hit procedural goals at comparable speeds, with crew 3 being slightly faster. Using resources for non-procedural operations thus did not slow down the progression through procedures.

Table 3: Crew size, duration of scenario and performance scores (OPAS measures and whether or not the team allocated resources to identifying the leak location) for scenario 2. Note that duration does not necessarily equate better performance.

Scenario 2 OPAS					
Crew	#1	#2	#3	#4	#5
Size	5	5	3	4	3
Attempt to identify?	Limited	Yes	No	No	Yes
Identification of leak in RCS	3	4	1,5	1,5	2
Identification of leak in RHR	3	4	1,5	1,5	4,5
Cooldown	4	4	4	3	1
Average	3.3	4.0	2.3	2.0	2.5
Duration	01:26:21	01:14:47	02:10:00	01:12:51*	01:30:37

*Scenario was stopped before the crew had completed the final goal

5.2.1. Factors Causing High Performance

The highest scoring crew was crew 2 with an average OPAS score of 4.0. Two aspects were observed that may have caused this performance: First, the crew was one of two crews (along with crew 5) that invested heavily into finding the location of the leaks. The crew decided early on that they needed information from external sourced information. Therefore, they communicated frequently with ‘Field Operators’ (roleplayed by the Control Room) throughout the scenario. Based on these communications, they were able to identify both leaks and to perform actions towards isolating them. Second, the crew’s Unit Supervisor chose not to read notes in the procedures aloud

Please cite as:
Andersen, E., Kozin, I., & Maier, A. (2018). Biased Decision Making in Realistic Extra-Procedural Nuclear Control Room Scenarios. In *Design Computing and Cognition DCC’18* (pp. 601-620). Springer. https://doi.org/10.1007/978-3-030-05363-5_30

and furthermore ended up skipping information in the procedures. In addition to these performance measures, the crew was furthermore the only crew to execute on restoring water to the Refuelling Water Storage Tank.

5.2.2. Factors Causing Medium Performance

The medium scoring crew was crew 1, with an average OPAS score of 3.3. One aspect was observed that may have caused this performance: Compared to two of the lower scoring crews, crew 4 discussed and performed some preliminary actions towards identifying and isolating the leak. However, rather than spending resources on investigating further the exact locations, they tried to deduce the locations from secondary information instead of testing their hypotheses with alternative sources of information, whereas crews 2 and 5 relied on dialogue with 'Field Operators'. Based on this information gathering, crew 1 was able to achieve some degree of identification of the location of the leaks, and to perform some isolating actions.

5.2.3. Factors Causing Low Performance

The lowest scoring crews were crew 3, 4 and 5, with OPAS scores of 2.3, 2.0 and 2.5 respectively. Different aspects were observed for each team that may have caused this lower performance: For crew 3 and 4, the lower score was caused by the fact that they did not attempt to identify and isolate the leaks. For crew 5, their lower score was caused by suboptimal performance during attempts at isolating the leak in the Reactor Coolant System and during cooldown

Crew 3 and 4 had preliminary suspicions and discussions about the possibility of a leak. Crew 3 decided early on that there was only one minor leak and to not spend resources on communicating with a Field Operator regarding alternative information – they only communicated with the Field Operator for practical tasks, such as energizing valves. Crew 4 discussed calling a FO to investigate further, but decided not to do so as they believed there to be too many possible candidates for the leak location.

A common factor for the lower scoring crews was that they did not show appropriate patience in watching the effects of the procedure actions. As a result, the crews entered subsequent procedures based on misleading information about whether or not the preceding procedure had been effective. Crew 5 attempted to compensate for this by re-running procedures while simultaneously entering another procedure. This increased the workload on the crew, which may in turn have caused the lower performance with regards to cooling.

As a final factor, crew 5 was, as in scenario 1, very active in calling for strategy briefs and in collaborative planning of future steps.

Please cite as:
Andersen, E., Kozin, I., & Maier, A. (2018). Biased Decision Making in Realistic Extra-Procedural Nuclear Control Room Scenarios. In *Design Computing and Cognition DCC'18* (pp. 601-620). Springer. https://doi.org/10.1007/978-3-030-05363-5_30

6. Discussion

This study investigated the decision making and performance of five crews of nuclear control room operators in two realistic simulated scenarios. The scenarios involved non-typical situations, which were caused by multiple failures in the nuclear power plant system, and could only be detected in full through deviation from procedures, such as autonomous requests or search for additional information. We found that some crews strictly adhered to procedures despite its leading to suboptimal performance. Furthermore, we found that crews did generally not persist in the mind-set of testing multiple hypotheses once procedures had been entered. This was seen in the reformulation and/or further commitment to a hypothesis that was consistent with the procedure in scenario#1, and a rationalization/explaining away of reasons to go outside of procedures in scenario#2. Crucially, these behavioural patterns were observed despite the fact that, generally, teams who also pursued actions outside of procedures were not slower or less accurate. Furthermore, we found examples of both types of bias behaviour. In the following, we elaborate our findings with respect to each bias and make suggestions for implications for design.

6.1. Expertise

Despite crews receiving extensive training in the simulated power plant, we observed several possible indicators of biased behaviour due to misapplied expertise: First, in scenario 2, the decision of crew 1, 3 and 4 to not allocate additional resources towards detecting the leaks could have been an optimal strategy at their home plant. Their expertise may have thus guided them to not continue on, as conserving resources would lead to greater success. Second, in scenario 2, crew 1 chose not to collect additional information from outside sources, as they believed they could reach sufficient data from secondary calculations in the control room, and crew 4 decided not to pursue identification because they believed there to be too many possible causes. These behavioural patterns are consistent with the tendency to not seek for additional information due to expertise, which, in this case caused suboptimal performance.

6.2. Confirmation Bias

6.2.1. Early hypotheses

Despite not being required to do so by the procedures, we observed that all crews created hypotheses about the cause of the problems early on in

Please cite as:
Andersen, E., Kozin, I., & Maier, A. (2018). Biased Decision Making in Realistic Extra-Procedural Nuclear Control Room Scenarios. In *Design Computing and Cognition DCC'18* (pp. 601-620). Springer. https://doi.org/10.1007/978-3-030-05363-5_30

both scenarios. Furthermore, in both scenarios, we found that for the majority of crews, the first hypotheses persisted through the entire scenario. Deviations were seen in scenario 1, where crew 5 continuously explored multiple hypotheses throughout the scenarios (thus never committing to one hypothesis) and where crew 3 committed to a hypothesis but reconsidered when until salient outside information from a field operator caused them to reconsider. In particular, as evidence by the performance of crew 4 in scenario 1, it seemed that the early beliefs of the Unit Supervisor were especially important for choice of strategy. These observations are thus consistent with the notion that operators were affected by confirmation bias in the form of commitment to early hypotheses

6.2.2. Confirmatory search

As expected, we found several examples of confirmatory search. In scenario 1, all but one crew had adopted the hypothesis that only one Steam Generator was damaged. To test this hypothesis in a non-confirmatory manner, crews would need to look for information about the integrity of the other steam generators and see whether there was damage in multiple locations. Furthermore, given that there was a considerable delay between Steam Generator #2 and #3 leaking, the hypothesis that only a single Steam Generator was damaged was true for an extended period of time. To avoid confirmatory search, crews would thus have to continuously look for changes in information that caused the previously true hypothesis to be false. We observed that only one crew, crew 5, employed a strategy that allowed them to continuously search for alternate sources of information, while another, crew 3, was given alternate information by the field operators, which prompted them to adopt the true hypothesis that two leaks had occurred. For the remaining crews, two sources of information were used to confirm the original hypothesis, which caused the teams to not search for additional information. First, after the rupture in Steam Generator#3 had reached its maximum, the severity of its effects was much larger. Consequently, crews could readily explain radiation measures from Steam Generator #1 and Steam Generator#2 as 'shine' effects, which is a common occurrence, and thus a theoretically valid data point to confirm that there was only damage to Steam Generator#3. Second, due to the relative small size of the Steam Generator#2 leak, it was difficult to detect the effects of the leak due to the presence of the large rupture in Steam Generator#3. In fact, looking at the instruments in the nuclear control room, the pressure was stable in Steam Generator#2 for extended periods of time. This could be interpreted as a valid data point for confirming that only Steam Generator#3 was damaged. Consistent with

Please cite as:
Andersen, E., Kozin, I., & Maier, A. (2018). Biased Decision Making in Realistic Extra-Procedural Nuclear Control Room Scenarios. In *Design Computing and Cognition DCC'18* (pp. 601-620). Springer. https://doi.org/10.1007/978-3-030-05363-5_30

our expectations, we found examples of both types of confirmatory search in the three remaining crews.

6.3. Implications for Design

Our results suggest that the presence of and design of the procedures may have been conducive to increased risk of being influenced by these biases. In scenario 1, we found that all crews made initial efforts to obtain local samples through communications with a field operator, but that all but one crew abandoned these efforts after emergency procedures were entered. In scenario 2, several crews specifically chose not to pursue any actions towards diagnosing the problem, as the procedures did not require doing so. These results thus suggest that the fact that procedures did not require diagnosis of the problem, nor encourage operators to look for alternate sources of information, had the opposite effect of dissuading crews from exploring alternate hypotheses.

While these findings are exploratory, rather than confirmed in a deductive way, previous studies have shown that interaction with written materials can cause a decrease in idea generation [18], [19], [38]. Furthermore, previous research has shown that confirmation bias also influences how pilots planning decisions in adverse weather conditions [39], and how military analysts prioritize information [40]. Therefore, the design of procedures and checklists that prevent (or minimize) these biases has great potential for improving performance and, thus, safety in these fields. To this end, research is needed that explores this relation. In the design field, such research has been conducted with some success with regards to minimization of design fixation [21]. Similarly, there have been attempts of debasing decision makers through design, albeit with mixed results [39], [40]. In particular, the use of counterfactuals, meaning examples that are opposite to the observed events or encourage thinking of the opposite of the present view, have been successful in combating confirmation bias[40], [41]. However, further research is needed to determine design interventions that could be successful for this purpose.

7. Conclusion

This paper presented a re-analysis of a data from two realistic simulation scenarios of a pressurized water reactor, which were conducted with five nuclear control room operator crews. The scenarios required operators to perform autonomous actions outside of procedures to achieve optimal performance. Drawing on literature from cognitive psychology and project management, we investigated whether the misapplication of expertise and

Please cite as:
Andersen, E., Kozin, I., & Maier, A. (2018). Biased Decision Making in Realistic Extra-Procedural Nuclear Control Room Scenarios. In *Design Computing and Cognition DCC'18* (pp. 601-620). Springer. https://doi.org/10.1007/978-3-030-05363-5_30

confirmation bias caused suboptimal performance in the applied case of a nuclear control room emergency. While the study presented is exploratory, and the findings have thus not been validated in a deductive manner, we found evidence that both biases could explain differences in performance. Furthermore, we found that the use of procedures may have increased the effect of confirmation bias. We concluded by discussing implications and opportunities for design of procedures.

Literature

[1] B. K. Burian, "Design Guidance for Emergency and Abnormal Checklists in Aviation," *Proc. Hum. Factors Ergon. Soc. 50 th Annu. Meet.*, vol. 50, no. October, pp. 1–6, 2006.

[2] B. K. Burian, "Do You Smell Smoke? Issues in the Design and Content of Checklists for Smoke, Fire, and Fumes," in *Emergency and Abnormal Situations Symposium*, 2005.

[3] N. Lau *et al.*, "Ecological interface design in the nuclear domain: An application to the secondary subsystems of a boiling water reactor plant simulator," *IEEE Trans. Nucl. Sci.*, vol. 55, no. 6, pp. 3579–3596, Dec. 2008.

[4] U. Weyer, A. O. Braseth, M. Eikås, L. Hurlen, P. Kristiansen, and J. Kvalem, "Safety Presentation in Large Screen Displays - A New Approach," in *SPE Intelligent Energy Conference and Exhibition*, 2010.

[5] A. O. Braseth and T. A. Øritsland, "Visualizing complex processes on large screen displays: Design principles based on the Information Rich Design concept," *Displays*, vol. 34, no. 3, pp. 215–222, 2013.

[6] J. Wreathall and J. Reason, "Latent Failures and Human Performance in Significant Operating Events," *JWCo. Inc*, 1993.

[7] U.S. Nuclear Regulatory Commission, "Review of Findings for Human Performance Contribution to Risk in Operating Events (NUREG/CR-6753, INEEL/EXT-01-01166)," 2002.

[8] S. Massaiu, "Critical features of emergency procedures: Empirical insights form simulations of nuclear power plant operation," *Training*, pp. 277–284, 2010.

[9] P. Du and E. F. MacDonald, "Products' Shared Visual Features Do Not Cancel in Consumer Decisions," *J. Mech. Des.*, vol. 137, no. 7, p. 71409, 2015.

[10] P. Du and E. F. MacDonald, "Eye-Tracking Data Predicts Importance of Product Features and Saliency of Size Change," *Vol. 5 25th Int. Conf. Des. Theory Methodol. ASME 2013 Power Transm. Gearing Conf.*, vol. 136, no. 8, p. V005T06A024, 2013.

[11] K. Goucher-Lambert and J. Cagan, "The Impact of Sustainability on Consumer Preference Judgments (DETC2014-34739)," *ASME 2014 Int. Des. Eng. Tech. Conf. Comput. Inf. Eng. Conf.*, vol. 137, no. c, pp. 1–11, Aug. 2014.

Please cite as:
Andersen, E., Kozin, I., & Maier, A. (2018). Biased Decision Making in Realistic Extra-Procedural Nuclear Control Room Scenarios. In *Design Computing and Cognition DCC'18* (pp. 601-620). Springer. https://doi.org/10.1007/978-3-030-05363-5_30

[12] K. Goucher-Lambert, J. Moss, and J. Cagan, "Inside the Mind: Using Neuroimaging to Understand Moral Product Preference Judgments Involving Sustainability," *J. Mech. Des.*, vol. 139, no. 4, p. 41103, 2017.
[13] S. J. Faerber and C. C. Carbon, "Jump on the innovator's train: Cognitive principles for creating appreciation in innovative product designs," *Res. Eng. Des.*, vol. 24, no. 3, pp. 313–319, 2013.
[14] T. N. Reid, E. F. MacDonald, and P. Du, "Impact of Product Design Representation on Customer Judgment," *J. Mech. Des.*, vol. 135, no. 9, p. 91008, 2013.
[15] S. T. Wei, L. C. Ou, M. R. Luo, and J. B. Hutchings, "Package design: Colour harmony and consumer expectations," *Int. J. Des.*, vol. 8, no. 1, pp. 109–126, 2014.
[16] M. Perez Mata, S. Ahmed-Kristensen, P. B. Brockhoff, and H. Yanagisawa, "Investigating the influence of product perception and geometric features," *Res. Eng. Des.*, vol. 28, no. 3, pp. 357–379, 2017.
[17] E. Andersen and A. Maier, "The attentional capture of colour in visual interface design: A controlled-environment study," in *Proceedings of the International Conference on Engineering Design, ICED*, 2017, vol. 8, no. DS87-8.
[18] D. G. Jansson and S. M. Smith, "Design fixation," *Des. Stud.*, vol. 12, no. 1, pp. 3–11, 1991.
[19] L. A. Vasconcelos and N. Crilly, "Inspiration and fixation: Questions, methods, findings, and challenges," *Des. Stud.*, vol. 42, pp. 1–32, 2016.
[20] J. B. Nikander, L. A. Liikkanen, and M. Laakso, "The preference effect in design concept evaluation," *Des. Stud.*, vol. 35, no. 5, pp. 473–499, Sep. 2014.
[21] V. K. Viswanathan and J. S. Linsey, "Design Fixation and Its Mitigation: A Study on the Role of Expertise," *J. Mech. Des.*, vol. 135, no. 5, p. 51008, Apr. 2013.
[22] D. Kahneman and G. Klein, "Conditions for intuitive expertise: A failure to disagree.," *Am. Psychol.*, vol. 64, no. 6, pp. 515–526, 2009.
[23] G. Gigerenzer and W. Gaissmaier, "Heuristic Decision Making," *Annu. Rev. Psychol.*, vol. 62, no. 1, pp. 451–482, 2011.
[24] S. Massaiu and L. Holmgren, "Diagnosis and Decision-Making with Emergency Operating Procedures in Non-typical Conditions: A HAMMLAB Study with U.S. Operators," *HWR-1121. Halden, Norw. OECD Halden React. Proj.*, 2014.
[25] S. Massaiu and L. Holmgren, "A HAMMLAB HRA Data Collection with U.S. Operators. HWR-1123. Halden, Norway: OECD Halden Reactor Project.," *HWR-1123. Halden, Norw. OECD Halden React. Proj.*, 2016.
[26] R. S. Nickerson, "Confirmation bias: A ubiquitous phenomenon in many guises.," *Rev. Gen. Psychol.*, vol. 2, no. 2, pp. 175–220, 1998.
[27] H. A. Simon, "What Is an Explanation of Behavior?," *Psychol. Sci.*, vol. 3, no. 3, pp. 150–161, 1992.
[28] J. Shanteau, "Competence in experts: The role of task characteristics," *Organ. Behav. Hum. Decis. Process.*, vol. 53, no. 2, pp. 252–266, Nov.

Please cite as:
Andersen, E., Kozin, I., & Maier, A. (2018). Biased Decision Making in Realistic Extra-Procedural Nuclear Control Room Scenarios. In *Design Computing and Cognition DCC'18* (pp. 601-620). Springer. https://doi.org/10.1007/978-3-030-05363-5_30

1992.

[29] J. Shanteau, “How much information does an expert use? Is it relevant?,” *Acta Psychol. (Amst).*, vol. 81, no. 1, pp. 75–86, 1992.

[30] S. E. Dreyfus, “The Five-Stage Model of Adult Skill Acquisition,” *Bull. Sci. Technol. Soc.*, vol. 24, no. 3, pp. 177–181, 2004.

[31] J. S. Bruner and M. C. Potter, “Interference in Visual Recognition,” *Science (80-.).*, vol. 144, no. 3617, pp. 424–425, Apr. 1964.

[32] L. Ross and C. A. Anderson, “Shortcomings in the attribution process: On the origins and maintenance of erroneous social assessments,” *Judgm. under Uncertain. Heuristics biases*, no. 1977, pp. 129–153, 1982.

[33] C. a. Anderson, M. R. Lepper, and L. Ross, “Perseverance of social theories: The role of explanation in the persistence of discredited information.,” *J. Pers. Soc. Psychol.*, vol. 39, no. 6, pp. 1037–1049, 1980.

[34] J. Baron, “Myside Bias in Thinking About Abortion,” *Think. Reason.*, vol. 1, no. 3, pp. 221–235, 1995.

[35] P. C. Wason, “On the failure to eliminate hypotheses in a conceptual task,” *Q. J. Exp. Psychol.*, vol. 12, no. 3, pp. 129–140, 1960.

[36] D. Griffin and A. Tversky, “The weighing of evidence and the determinants of confidence,” *Cogn. Psychol.*, vol. 24, no. 3, pp. 411–435, 1992.

[37] G. Skraaning Jr., “The operator performance assessment system (OPAS, HWR-538),” 1998.

[38] L. A. Vasconcelos, C. C. Cardoso, M. Sääksjärvi, C.-C. Chen, and N. Crilly, “Inspiration and Fixation: The Influences of Example Designs and System Properties in Idea Generation,” *J. Mech. Des.*, vol. 139, no. 3, p. 31101, 2017.

[39] S. Walmsley and A. Gilbey, “Debiasing visual pilots’ weather-related decision making,” *Appl. Ergon.*, vol. 65, pp. 200–208, 2017.

[40] M. B. Cook and H. S. Smallman, “Human Factors of the Confirmation Bias in Intelligence Analysis: Decision Support From Graphical Evidence Landscapes,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 50, no. 5, pp. 745–754, 2008.

[41] A. D. Galinsky and G. B. Moskowitz, “Counterfactuals as Behavioral Primes: Priming the Simulation Heuristic and Consideration of Alternatives,” *J. Exp. Soc. Psychol.*, vol. 36, no. 4, pp. 384–409, 2000.

Please cite as:
Andersen, E., Kozin, I., & Maier, A. (2018). Biased Decision Making in Realistic Extra-Procedural Nuclear Control Room Scenarios. In *Design Computing and Cognition DCC’18* (pp. 601-620). Springer. https://doi.org/10.1007/978-3-030-05363-5_30