Classifying Risk Uncertainty for Decision Making

Dan Port
University of Hawaii
Associate Professor
Department of Information Technology
Honolulu, USA
dport@hawaii.edu

Joel Wilf
University of Hawaii
PhD Student
Communications and Information Sciences (CIS)
Honolulu, USA
jwilf@hawaii.edu

Abstract
Studies of NASA mishaps often reveal a flawed decision-making process — one that underestimates risk. In this paper we turn our attention from the risk itself to uncertainty about the risk. In particular, we look at how decision-making accounts for uncertainties about a risk’s likelihood of occurring and the consequence if it does occur. We propose a simple way of classifying risks according to these uncertainties. Then we use this classification scheme to gain insight into the flawed decision-making that contributed to the Challenger disaster and other NASA mishaps as well. We show how our risk classification scheme can improve decision-making and help avoid mishaps in the future.

1. Introduction

In our study of NASA mishaps, we found a recurring pattern of flawed decision-making [1]. Decision-makers often mis-estimated one or more risks to the mission. There were many reasons for this error, including overconfidence, cognitive bias, and groupthink [1]. The results were poor quality decisions that, in some cases, led to the loss of missions and even human life [1], [2], [3].

In this paper, we turn our attention from how decision-makers mis-estimate risk to how they fail to account for uncertainty in risk parameters. Traditionally, risk is defined as:

\[
\text{Risk} = (\text{Likelihood of Occurrence}) \times (\text{Consequence})
\]

Where Likelihood is typically expressed as a probability (i.e., a number between 0 and 1), and Consequence is the magnitude of the loss experienced if the risk is realized (often expressed as a dollar amount) [4].

In this paper, we examine the case where decision-makers estimate the Likelihood and Consequence, but don’t take into account the uncertainties embodied in these estimates. We show that incorrectly handling the uncertainties can lead to poor quality decisions. We take examples from the domain of earth and space exploration to show how this has contributed to mission failure.

We then propose a mitigation for this problem: a simple scheme for classifying risks that gives decision-makers a way to think about and incorporate important information about risk uncertainties into their decisions.

The paper is organized as follows. In Section 2, we will describe the motivation for our study and why we chose to examine data from NASA mishaps. In Section 3, we look briefly at related work and approaches taken in the past. In Section 4, we examine the fateful decision to launch the Challenger Space Shuttle in detail, as an example of decision-makers not accounting for Likelihood uncertainties. We then briefly survey other NASA mishaps where similar mishandling of risk uncertainties may have contributed to mission failure. In Section 5, we propose a simple model for classifying risks according to the uncertainties in Likelihood and Consequence. In Section 6, we show how to use this model in the decision-making process. In Section 7, we show how this approach could have been applied in our Challenger example. In Section 8, we briefly examine how our model could be applied to other NASA mishaps. Finally, in Section 9, we conclude with a summary of the research contribution, the limitations of our study, and the potential for future work.

2. Background and Context

Our examination of decision-making, risk, and uncertainty is motivated by trying to improve the odds of success for earth and space missions. In NASA tradition, this is often accomplished by analyzing past failures and deriving lessons learned [5].

This particular study started from our personal observations of how risks are handled in the development of earth and space science missions. As mentioned earlier, we noticed that risks are often incorporated into decision-making, but uncertainties about those risks are often not considered.

The reason is that risk is often seen as falling in one of two categories: 1) as “known” risks, where there is
no need to account for uncertainty or 2) as “unknown” risks, where there is no ability to account for the uncertainty. In both cases, the uncertainty is not directly used in the decision calculation itself.

We have experienced review meetings where risks were identified, discussed, and assessed – after which, the reviewers credited themselves for doing risk-informed decision making – and adjourned.

We have also witnessed what happens when decision-makers disagree on risk assessment. Sometimes this led to an arbitrary decision (e.g., “splitting the difference” between the two opinions for the sake of concession). On other occasions, it led to ignoring the risk in the decision altogether, because it was “unknowable”.

We questioned whether these decision practices took place on NASA missions that experienced mishaps, whether these practices contributed to the mishaps, and if so, what could be done about it. We sought the answers in focused study of NASA mishap reports and detailed studies, as are available for the largest failures, such as Challenger and Columbia [6], [7].

3. Related work

There is an extensive literature on risk and uncertainty in decision making. Of interest for this study are those that discuss schemes for classifying uncertainty that can be used in decision making. For example, one scheme of classifying uncertainty comes from the economist Frank Knight, who distinguished between uncertainty for which a probability distribution is known and that for which one is unknown or unmeasurable (“Knightian Uncertainty”) [8].

Another, albeit informal, scheme for classifying uncertainty involves what we know and “what we know we know” (e.g., things we know we know, things we know we don’t know, things we don’t know we don’t know) [9]. Oddly, we have found many discussions of this four-quadrant scheme, though the approach is more philosophical than for practical use in decision-making [10]. Our classification scheme, described below in Section 5, bears a surface resemblance to this typology, but it is designed specifically for use in decision analysis.

Since we are using examples from the domain of NASA missions, it is important to understand NASA’s standards for risk and decision making. Two key documents are the Agency Risk Management Procedural Requirements [11] and the NASA Risk Informed Decision Making (RIDM) Handbook [12], which offers more detailed guidance. The RIDM handbook reminds decision makers that “decisions are made taking into account applicable risks and uncertainties” and that “complete assessment of [risk] likelihood also calls for characterization of its uncertainty” [12]. This is sound advice, and the sections below will provide additional details on motivation (through mishap analysis) and method (through a proposed model of risk uncertainty).

In addition, this study makes use of NASA mishap reports [13], [14], [15], [16], [17], [18], [19], [20]. It also draws on some of the larger, independent studies of the tragic Challenger and Columbia shuttle accidents [2], [3], [21], [22]. Decisions that contribute to mishaps are sometimes well-described, sometimes inferred in the smaller reports, whereas decision-making is an important theme in the major works on the shuttle disasters. This is especially true for the Challenger launch decision [2].

4. The Challenger Example

At the heart of the Challenger launch decision was uncertainty about the risk that O-rings – rubber seals on the Challenger’s fuel tank – would fail catastrophically. The O-rings in the booster rockets on the space shuttle were designed to expand when heated, in order to seal different chambers of the rocket, so that the solid rocket fuel would not be ignited prematurely. According to engineering specifications, the O-rings must expand by some amount, say at least 5%, to ensure a safe launch. When an O-ring does not expand by at least this amount there is a risk of a “blowout” failure where fuel leaks out and may ignite outside the booster shell and very likely to cascade into an explosion. This was estimated to occur 1 out of 1000 times if the expansion is less than 5% [23].

O-ring degradation during flight was a known risk. What was unknown was the relationship between temperature and O-ring degradation. The scatterplot, below, shows data on the number of O-rings that failed to expand more than 5% (an “incident”) and the temperature at launch [23].

![Figure 1 – O-Ring failures vs. temperature](image_url)
There were only 24 data points, and none that extended to the temperature on the launch-pad on the morning of the launch. Therefore, the decision makers had to extrapolate the results for an as-yet untested temperature. They used the linear regression curve shown as the diagonal line in Figure 1. The following analysis is only loosely based on the account of what transpired. We describe a variation on the analysis performed to better illustrate the risk-classification issues involved. However, the resulting misclassification of risk would be the same if we followed the analysis actually performed.

According to the linear regression model, there is a highly significant negative linear relationship between temperature and number of incidents (p-value less than 0.001). At the time of the launch decision the ambient temperature was 9 degrees Fahrenheit, giving a 95% prediction interval of between 4 and 18 O-ring incidents– would be expected. From a safety perspective, they might have even used a 99.99999% confidence level and calculated a extremely confident worst case of 38 O-ring incidents [24].

This leads to an “extremely confident worst case” failure probability of:

\[ P = \text{one blowout per 38 incidents} = 0.03634 \]

This is less than 4%, which given their 5% threshold, placed them within their margin of safety. Under this analysis, the decision to launch the Challenger is justified.

However, their analysis does not fully account for the uncertainty about the relationship between O-ring incidents and low temperatures, for which measurements weren’t available. To obtain the number of incidents at the low temperatures, they relied on linear regression. The uncertainty not accounted for is the uncertainty in linear extrapolation to a temperature well outside the range supported by the data.

In fact, we easily see that a linear model is not appropriate: a normal Q-Q plot of residuals (a plot of the differences between the observed and predicted values) should not produce a high, outlier value (e.g., greater than 3 std errs). Figure 2 shows the Q-Q plot for this linear regression:

\[ N = aT^b \]

Where \( N \) is the number of O-ring incidents and \( a \) and \( b \) are constants determined by the best fit to the data. Under this model, temperature was also highly significant (p-value about 0.001) and no residual outliers. We calculated \( N=113 \) expected incidents at 29 degrees giving a probability of a blowout of 0.1061, or about 11% -- well over the safety margin! This does not account for the variability, especially from extrapolation. Considering the 95% prediction interval here would give a worst-case of 1689 incidents with a .82 probability of a blow out! In this characterization of the uncertainty in extrapolation the decision would decidedly be not to launch. The point to consider here is, given the uncertainty in how to appropriately extrapolate to a temperature not supported by the data, why did they settle on the clearly more optimistic assessment of blowout risk? Feynman in the Challenger disaster result suggested it was a case of acclimation bias [22]. By better understanding the source of uncertainty and how it affects decision making, such biases perhaps could be reduced.

It’s interesting to ask how low the temperature could go for the result to be within the 5% safety margin. Figure 3 is a plot of the predicted number of blowouts (y-axis) at each temperature (x-axis). The solid black
The lower red curve shows the most likely number and the upper red curve shows the maximum number within a 95% confidence interval. For reference, the lower red curve shows where the 95% confidence range fell in the linear model actually used by the decision makers. The 5% safety margin is indicated by the dotted line:

![Figure 3 – Blowouts vs. temperature](Image)

Figure 3 shows that the number of predicted blowouts crosses the safety-margin at about 35 degrees, and the curve representing 95% confidence crosses at about 49 degrees. The difference from the linear model is striking. At the actual temperature of 29 degrees, the decision using the power model would be “don’t launch!”

As mentioned previously this discussion is not an account of the analysis performed. Apparently, they did use a simple linear model, but only using data when there was an O-ring incident. Their regression analysis conclusion was there was a significant relationship between temperature and increased O-ring incidents. They extrapolated by assuming there is no relationship at lower temperatures outside the range of data for the model and concluding that the simple 1 in a 1000 chance of blowout was that risk was acceptable to launch. But the critical point is not the different results given by different models. It’s in how the models were compensating for uncertainty that was not acknowledged by the decision makers. The models were used to extrapolate from a small sample of measurements, to estimate the likelihood of a catastrophic blowout at the real temperature, which lay outside the range of their sample.

The decision makers had no O-ring test data to support the confident expectation of a successful launch in such cold conditions. Engineers who worked on the Shuttle delivered a biting analysis: “We’re only qualified to 40° F. No one was even thinking of 18° F. We were in no man’s land.” [24] The Rogers Commission did an extensive review of the Challenger disaster. They found NASA’s decision-making processes among the contributing factors to the accident [6]. In retrospect, there was plenty of understanding and discussion of the uncertainties present and the risk they presented.

The question for us is how to model such risk uncertainties, to bring them out of the shadows so to speak, to make them part of the decision-making process, as should have been done with the uncertainties surrounding the O-ring risk on the Challenger.

### 5. Modeling Risk Uncertainties

Risk has two dimensions: Likelihood and Consequence. Therefore, the simplest way to account for risk uncertainty is to determine whether 1) the risk’s Likelihood is known or unknown and 2) whether the risk’s Consequence is known or unknown. Since Likelihood is a probability measure, a known Likelihood means being able to assign a probability to it. A known consequence means being able to assign the magnitude of a loss to it (e.g., assign a dollar value for its cost).

Thus, we define a classification quadrant of (known Likelihood, known Consequence), (known Likelihood, unknown Consequence), (unknown Likelihood, known Consequence), and (unknown Likelihood, unknown Consequence), abbreviated as KK, KU, UK, and UU as shown in Figure 4:

![Figure 4 – Risk uncertainty quadrant](Image)

To illustrate this, consider how we would classify the risk of betting on the flip of a coin:
1. **KK** – if the coin is fair and the bet is $1 in advance
2. **UK** – if the coin has an unknown bias and the bet is $1 in advance
3. **KU** – if the coin is fair, but the bet is set by the opponent while the coin is in mid-flip
4. **UU** – if the coin has an unknown bias, and the bet is set by the opponent while the coin is in mid-flip

As you can see, this method makes the uncertainty in risk factors explicit. The main reason for doing this is to address a second-order risk -- or “decision risk” -- that appears when there is uncertainty in the risk factors. Decision risk relates to how confident we are that the decision is the correct one. Without taking decision risk into account, it is almost inevitable that decision-makers will fall prey to natural biases and become overconfident in the correctness of the decision [25].

For example, if we didn’t know there was a bias in a coin flip we were betting on, we would assume it was a fair 50-50 bet. In fact, what we are doing is treating an unknown known (UK) as a known known (KK). The result is that we are incorrectly assessing our risk: we are overconfident in a decision made using an incorrect assumption. We are taking on more risk than we accounted for, since we are not accounting for decision risk.

In the next section, we discuss how to use this model to recognize and incorporate decision risk into the decision process.

6. Using Classification in the Decision Process

The benefit of our risk classification model is that the strategy for handling decision risk is different for each of our four quadrants. If the risk uncertainty is KK, the expected outcome of the decision is simply Likelihood x Consequence, a multiplication of two known quantities. The decision strategy would be the classic one, in which you can compute the expected benefit or loss for each decision option. In our coin flip example, the expected risk exposure for both heads and tails is $1 \times 0.5 \text{ (even odds)} = $0.50.

In the other three quadrants, we are confronting uncertainty and hence, decision risk. Classifying risk as UK, KU, or UU guides you towards assessing the uncertainty you have in either (or both) Likelihood or Consequence, and determining how this uncertainty factors into the decision.

One of the implications of classification is that decision-makers need to consider their risk posture before they can analyze the risk. They need to make it explicit whether they are primarily benefit-seeking or risk-avoiding. If risk-avoiding, then they need to determine how uncertainty affects the worst case for risk. If benefit-seeking, then the decision makers need to determine how uncertainty affects the best case. What is the upper bound on the expected loss? What is the lower bound on the expected gain? Can we calculate a Value at Risk (VaR) for the decision? That is, we should be able to assert, “We have a 95% confidence level that our loss will not exceed this value.”

Classifying risk as UK, KU, or UU opens up the possibility of an additional option to consider: the best choice may be to reduce the uncertainty first, before making the final decision. Decision-makers often don’t consider this option. As we’ll show in Section 8, below, we found this to be a factor in several NASA mishaps.

Reducing uncertainty typically takes different forms when dealing with Likelihood (UK/UU) or Consequence (KK/UU). Likelihood uncertainties are usually reduced by obtaining more information, for example, by running more tests or performing more detailed analysis. Consequence uncertainties may involve intervening in the system itself, for example creating redundancy to lower the impact of losing a critical sub-system. Buying insurance, where possible, is another strategy for reducing uncertainty in Consequence.

Let’s look at the KK, UK, and UU strategies for our example of betting on the flip of a coin. For unknown Likelihood (UK/UU), we might lower our bet for a number of rounds while we learn more about possible bias in the coin, using Bayesian statistics. For unknown Consequence (KU/UU) we may negotiate with our opponent to put a limit on the size of the bet – or pay a third party to “insure” against our losses.

This brings us back to the Challenger launch decision. How did the decision-makers handle the uncertainty? How should they have classified the risk? And what difference could it have made?

7. Challenger Revisited – with the Model

In terms of our risk classification scheme, the Challenger decision-makers treated the O-ring risk as though it were KK, known Likelihood and Consequence.

We agree that the Consequence was known. Everyone involved with the launch decision knew that an O-ring blowout would lead to a catastrophic failure. Burning gases would ignite the fuel tank and cause the horrific explosion we saw when Challenger launched.

However, we think the O-ring risk should be classified as UK. The decision-makers were not
accounting for uncertainty in the Likelihood. As we discussed in Section 4, they didn’t have test data for the actual temperature during launch. The data they did have could be extrapolated using various models that gave conflicting results (differences in the linear vs. power-law model discussed above).

Treating the risk as UK might have made a difference in the decision. Being conscious of the uncertainty, the engineers may have run their extrapolation through the more conservative, power-law model, which would have made a stronger case for a no-launch decision.

But more importantly, it might have led the decision-makers to the strategy for UK risks that we proposed in Section 5. It would have guided them to consider the worst-case risk, given the uncertainty, and then to delay the launch and buy down the uncertainty with more testing the O-rings at low temperatures.

8. Other Mishap Examples

Challenger is an excellent example of how misclassifying risk uncertainty can lead to bad decision-making and mission failure. But it’s important to verify that this pattern applies to a broader range of mission mishaps.

In an earlier study, we examined a set of ten NASA mishaps, taking place from 1986 to 2013 and ranging in impact from the loss of life down to being only a “close call” [1]. These are shown in Table 1:

<table>
<thead>
<tr>
<th>Year</th>
<th>Mishap</th>
<th>Mission</th>
<th>Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>1986</td>
<td>Challenger</td>
<td>Shuttle</td>
<td>Life</td>
</tr>
<tr>
<td>1990</td>
<td>Hubble Telescope</td>
<td>Telescope</td>
<td>Recover</td>
</tr>
<tr>
<td>1999</td>
<td>Mars Climate Orbiter</td>
<td>Orbiter</td>
<td>Mission</td>
</tr>
<tr>
<td>1999</td>
<td>Mars Polar Lander</td>
<td>Lander</td>
<td>Mission</td>
</tr>
<tr>
<td>2003</td>
<td>Columbia</td>
<td>Shuttle</td>
<td>Life</td>
</tr>
<tr>
<td>2004</td>
<td>Genesis</td>
<td>Sample Return</td>
<td>Recover</td>
</tr>
<tr>
<td>2004</td>
<td>Helios Drone</td>
<td>Drone</td>
<td>Mission</td>
</tr>
<tr>
<td>2005</td>
<td>Dart Earth Orbiter</td>
<td>Earth Orbiter</td>
<td>Mission</td>
</tr>
<tr>
<td>2010</td>
<td>Compton Telescope</td>
<td>Balloon</td>
<td>Mission</td>
</tr>
<tr>
<td>2013</td>
<td>Helmet Water</td>
<td>Space walk</td>
<td>Close</td>
</tr>
</tbody>
</table>

Table 1 – Set of NASA Mishaps (1986-2013)

Reviewing these mishaps, we found that for over half of them, an uncertainty surrounding an important risk was mis-classified – and this contributed to the mishap. For example, the Columbia shuttle mission handled the risk of its fuel-tank foam insulation falling off as a known known (KK). They thought the consequences were known: they knew foam fell off during testing and even other shuttle flights, without damaging results. In fact, this risk should have been classified as known unknown (KU), since the consequences were uncertain. Foam falling off and striking the shuttle at high speed ultimately led to its destruction [7].

We saw a similar pattern in the “Helmet Water” close call. Mission operations treated the risk of a leaky tube in the astronaut’s helmet as a known known (KK), since the tubes were known to leak on occasion and the consequence had so-far been minimal. In fact, the risk should have been classified known unknown (KU). There was uncertainty in the consequence, and during one space-walk, leaks from the tube caused an astronaut’s helmet to fill up with water. Fortunately, he made it safely back to the shuttle – and thus the mishap was a “close call” [20]

In another example, developers of the Hubble space telescope mirror seemed only aware of known known (KK) risks. However, uncertainties about the quality of the Hubble mirror were actually in the unknown unknown (UU) category. This would have suggested the strategy of "buying-down" the uncertainty through additional quality assurance, which in turn might have led to the discovery of the mission-crippling Hubble mirror flaw before its launch [13].

A similar classification and change in decision strategy also seemed to apply to the Mars Climate Orbiter (MCO). The project knew there was something amiss with the spacecraft trajectory. Treating it as an unknown, unknown (UU) risk and buying down the uncertainty with additional tests might have helped them find the cause – mixing “miles” and “kilometers” in their navigation calculations -- prior to the failed orbit insertion maneuver [14].

9. Conclusion

In this study, we suggested the importance for decision-makers to properly account for the uncertainties surrounding risks. We analyzed the Challenger launch decision as an example, showing how the decision-makers mis-handled uncertainty in the Likelihood of an O-ring blowout at low temperatures, and how this led to a flawed decision to launch. We showed that a similar pattern could be found in other NASA mishaps, as well.

We proposed a new risk classification scheme based on uncertainties about risk Likelihood and Consequence; and we presented decision strategies for each of our four risk uncertainty classifications. We applied our classification scheme and the resulting decision strategy to the Challenger launch decision and discussed its application to other NASA mishaps.

8.1. Contribution and Future Work
The contributions of this research include: a new framework for classifying risk uncertainties, strategies for using the classification scheme, and its application to decision-making under risk and uncertainty. This research also contributed to a greater understanding of the decision-making errors that contributed to the Challenger disaster and other NASA mishaps.

Future work includes a broader application of this framework to decision-making in other high-risk domains, for example medicine or public safety.

Another potential application is to risk management standards and practices. For example, the NASA Risk Informed Decision Making Handbook advises mission decision-makers to “take uncertainty under consideration” – but doesn’t provide a method for doing so [12]. Using the framework and decision strategies presented here, risk standards could offer concrete guidance in this area.

Ultimately, the goal of this and related research remains one of learning the lessons of past mishaps, adjusting our standards and practices accordingly, and steadily improving our odds of mission success.

9. References


