Near-Miss Evaluation Bias as an Obstacle to Organizational Learning: Lessons from NASA

Robin L. Dillon
McDonough School of Business
Georgetown University
Washington DC 20057
202-687-5398
rld9@georgetown.edu

Catherine H. Tinsley
McDonough School of Business
Georgetown University
Washington DC 20057
202-687-2524
tinsleyc@georgetown.edu

and

Edward W. Rogers
NASA Goddard Space Flight Center
NASA/GSFC, Code 170
Greenbelt, MD 20771
301-286-4467
edward.w.rogers@nasa.gov

Near-Miss Evaluation Bias as an Obstacle to Organizational Learning: Lessons from NASA

Abstract

After the Shuttle Columbia catastrophe, the investigation board (CAIB) stated that NASA needs to develop a "learning culture", meaning a capability to learn from past failures by understanding the technical and organizational causes of these mistakes (CAIB report, 2003). While many organizations learn from obvious failures, we argue that it is harder for organizations to learn from near-miss events (i.e., situations where a failure does not occur but nearly did), because these near-misses are processed as successes. For the shuttle program, prior debris problems could have caused a similar failure as on the Columbia mission except that the large pieces missed the highly sensitive portions of the orbiter. This acceptance of foam debris was adopted as a normal occurrence by the shuttle program managers similar to the problems at the time of the Challenger Disaster (detailed in Vaughan, 1996). We extend that work to show that an outcome bias influences people's evaluation of project managers, such that managers of failed missions were perceived more poorly than managers who made the same decisions but whose mission ended in either success or a near-miss. The similarity of ratings between the near-miss and success condition imply that even when a problem occurs that is clearly linked to prior managerial decisions, if the project is not harmed because of good luck, that manager is not held accountable for faulty decision making and neither the individual manager nor the organization learn from the experience potentially increasing the likelihood of a failure in the future.

Keywords: Near-Miss Bias, Decision Making, Organizational Learning

Near-Miss Evaluation Bias as an Obstacle to Organizational Learning: Lessons from NASA

Failures when examined with hindsight generally are attributed to poor decision making. As Fischoff (1982) explains, once outcomes are known, people tend to anchor on these outcomes and exaggerate what could have been anticipated at the time of the decision, even misremembering their own predictions to be consistent with now-known outcomes (Fischoff & Beyth, 1975). That people anchor on outcomes and how the valence of the outcomes play an overly influential role in decision evaluations has come to be known as an "outcome bias". Specifically, an outcome bias occurs when the outcome of a decision, rather than the appropriateness of the decision, systematically influences people's evaluations of the quality of the decision making (cf- Baron & Hershey, 1988; Allison, Mackie, & Messick, 1996). Yet, sometimes outcome failures are simply the result of bad luck, and sometimes successes are simply good luck. Consider a California community preparing for possible earthquakes. The community can make the same preparation decisions, but experience completely different outcomes if an earthquake occurs on a Friday afternoon during rush hour (i.e., Loma Prieta in 1989) versus at 4:30 am on a holiday (i.e., Northridge in 1994). In the case of a spacecraft mission to Mars, bad luck would include problems caused by terrain features (i.e., hitting a sharp rock), severe weather, or launch failures. In the case of the Columbia accident, the catastrophe was partly caused by poor decision-making and partly by bad luck. The shuttle experienced bad luck in that the piece of foam debris struck the leading edge of the wing, a very sensitive portion of the orbiter. However, poor decision-making by project managers allowed similar pieces of foam to become detached at least thirty times on previous missions and only good luck prevented a large enough piece of foam from striking a highly sensitive area on any of these previous

shuttle missions (CAIB, 2003). A critical question from the point of organizational learning is thus to try to isolate the extent to which any organizational outcome is primarily a result of decision making and when or how much is a result of luck.

In this paper we specifically explore how to more accurately deconstruct a decision outcome into the components of decision quality and luck by isolating a possible near-miss bias. A near-miss is an event that has some probability of a negative (even fatal) outcome and some probability of a positive (safe) outcome, but the actual outcome is non-hazardous. It is a "miss" in that the outcome is non-hazardous; it is "near" in the sense that the hazardous or fatal outcome could have occurred (Tinsley & Dillon, 2005; Dillon & Tinsley, 2005). A miss is the non-hazardous or successful organizational outcome and a hit would be a hazardous or failed organizational outcome. A near-miss event is a success that could have been a failure except for good luck. Prior research on near-miss events documents that the luck component of the near-miss is discounted (Tinsley & Dillon, 2005; Dillon & Tinsley, 2005). Assuming luck is discounted, then near-misses, rather than being coded as almost hits will be coded as almost successes. Thus, near-miss events will not be attended to as they should; managers will not be held accountable for near-miss events, and as a result, organizations will fail to learn what they might from these experiences. This challenge seems to apply to both public and private organizations. We have begun our research in the highly visible arena of public space exploration because of the clear specificity of outcomes and the intense scrutiny with which failures are analyzed while near-misses and successes are rarely analyzed. Furthermore, NASA has specifically been called upon to function more like a learning organization in its future endeavors.

The Outcome Bias in Managerial Evaluations

According to rational decision theories (e.g. Dawes, 1988), decision makers should be evaluated based on the appropriateness of their decision given their information at the time, rather than the consequences of that decision. Because all decisions are made under uncertainty, Edwards and colleagues liken a decision to a bet and chasten that evaluating a decision as "good or not must depend on the stakes and the odds, not on the outcome" (Edwards, Kiss, Majone, & Toda, 1984, 7). Yet, in organizational settings it can be difficult to judge decision quality, especially as ambiguity of the situation or novelty of the decision task increase. Agency theorists have argued extensively that organizations have difficulty monitoring managerial actions (Jensen & Meckling, 1976). Thus, it would be reasonable to assume organizations have difficulty judging the quality of a managerial action such as decision making. Because of this inability to monitor activities like managerial decision making, the consequential organizational outcomes of these decisions, and the valence of these outcomes, are used as proxies for managerial decision making. Hence, in organizations, as managers are evaluated for their decision making, we should see evidence of the outcome bias. Managerial decisions are likely to be evaluated based on the valence of the outcomes they produce, rather than on any "rational" analysis of the quality of the decisions themselves.

This outcome bias in evaluating managerial action in organizations suggests that when a manager engages in a set of decisions that result in an organizational success that manager's decision making should be evaluated more favorably than a manager whose set of decisions result in organizational failure—even if both managers made the exact same set of decisions. Given the general need for cognitive consistency, inflated ratings of a manager's decision ability (in the case of organizational successes rather than failures), may also produce significantly

higher ratings of a manager's other characteristics, such as competence, intelligence, leadership ability, and general promotability. This would be consistent with Baron and Hershey's (1988) data (study 4), showing that outcome bias extended beyond decisions to personal evaluations. Although none of Baron and Hershey's studies were in an organizational context (they studied mostly student's evaluations of others' gambling decisions), we expect a similar spill-over effect to occur in our organizational setting, as well, because of an evaluators' need for coherence in their judgments of others (cf- Festinger, 1957).

In our terminology, an organizational failure is called a "hit", in that the organization was hit by a failure or a disaster. Thus, an organizational success is called a "miss".

Hypothesis 1a: Managers whose decisions result in a miss (organizational success) will have their decision making evaluated in a significantly more favorable light than managers whose decisions result in a hit (organizational failure).

Hypothesis 1b: Managers whose decisions result in a miss (organizational success) will be judged to be more competent, to be more intelligent, to have more leadership ability, and to be more promotable than managers whose decisions result in a hit (organizational failure).

The Near-Miss Effect

Near-misses are events that could have ended in organizational failure but that negative outcome was narrowly avoided. Narrowly avoiding a hit implies some risk of a future hit and some component of luck that the current event was not a hit (but rather a near-miss). We think it is useful to look at how these organizational outcomes (and thus the decisions preceding these outcomes) are judged as evidence of potential near-miss bias. On the one hand, a near-miss

could be celebrated as a success (a miss); evidence of a system's resilience as failure is avoided. On the other hand, a near-miss could be soberly evaluated as a failure (a hit); evidence of a system's vulnerability as a risk was taken in ignorance and failure was narrowly avoided. We first consider, prescriptively, how near-miss events should be judged, and then consider descriptively how near-miss events are most likely to be judged. Finally, we consider the organizational learning that occurs in the alternative scenarios.

Prescriptively, of course the "right" answer for how to interpret a near-miss event (as an almost success versus an almost failure) depends on the "true" risk involved in whatever organizational system is producing the outcome. If the "true" risk is high, then the near-miss should be coded as an almost failure in that the organization attends to the information from the near-miss event and learns how to take steps to decrease the risk in the future. If, on the other hand, the "true" risk is low, then the near-miss event should be coded as an almost success so that the organization learns that the status quo is appropriate. The problem is that to divine the "true" risk requires objectively processing the outcome events, which means accurately categorizing them as either some form of miss or some form of hit, and accurately assessing the extent to which luck played a role in the actual outcome. Hence, the question poses an endogeneity problem: we need to know the true risk to know how to categorize a current near-miss event, but to know the true risk requires having properly categorized prior near-miss events.

So how, descriptively do people get out of this circularity? Do they judge a near-miss event positively as an almost-miss or negatively as an almost-hit? Prospect theory suggests, an outcome's valence is not judged in isolation but relative to a reference point (Kahneman & Tversky, 1979). So the valence of a near-miss event should be judged good or bad, depending on the reference point to which it is compared.

Norm theory (Kahneman & Miller, 1986) suggests that a stimulus event or outcome evokes its own frame of reference for evaluation by evoking its own normative alternatives. Normative alternatives here refer to what could, should, or might have occurred, but did not. This is what others have called counterfactual thinking (Galinsky et al., 2000). According to norm theory, dominant features of the stimulus event, such as the amount of risk involved, are said to be immutable and to guide people's spontaneous search for appropriate comparison alternatives. Those features of the event that could easily be imagined to have occurred differently are said to be mutable, and are what differentiate the normative alternative from the event itself. The more easily the mutable features can be imagined to occur, the more "close" is the normative alternative to the event itself. Closer alternatives evoke stronger judgments (as to an outcome's valence) than less close alternatives (Miller & MacFarland, 1986).

So what normative alternatives are evoked for a near-miss event? If a near-miss event evokes an organizational success as the comparison alternative, this should decrease the valence of the near-miss event. The thought process would be that this near-miss event was almost, but was not, a complete success. This near-miss event is less desirable than a success (a miss), which should generally lower the valence rating of the near-miss event. If, on the other hand the near-miss event evokes an organizational failure as the comparison alternative, this should increase the valence of the near-miss event. The thought process would be that this near-miss event was almost, but was not, an organizational failure. The near-miss event would be more desirable than a failure (hit), which should generally raise the valence rating of the near-miss event.

Neither norm theory nor prospect theory makes any restrictions on the number of normative alternatives that can be evoked as reference points. If both normative alternatives are

evoked, then near-miss outcomes should be evaluated as more positive than hits but more negative than misses. Thus, managers whose decisions result in a near-miss should be judged between those whose decisions end in hits and those whose decisions end in misses. However, one normative alternative may be closer to a near-miss than the other normative alternative, and if so, then it should exert more of an effect on the valence of the near-miss outcome.

We propose that when a near-miss occurs it is easier to imagine how that near-miss event was different from a hit (failure) than it is different from a miss (success). Said otherwise, a near-miss is more like a miss than it is a hit. Both near-misses and misses end in success, rather than failure—even though the near-miss was almost a failure. Therefore, the hit outcome (failure) becomes a more salient alternative outcome state than the miss outcome, reinforcing that this near-miss was not a hit, and enhancing the valence of the near-miss outcome. Hence, we expect the near-miss outcomes to be judged more similarly to the miss outcomes than to the hit outcomes.

This result would be consistent with prior research on the near-miss bias (cf- Tinsley & Dillon, 2005), which showed that people discounted the luck component in processing a near-miss event. When participants personally experienced near-miss events they were more risky in subsequent decision making tasks than people who did not experience near-miss events. This subsequent risky decision making suggests participants who experienced near-miss events were discounting their own good luck because the near-miss was categorized as a miss.

While prior research has shown people discount their own good luck in experiencing near-miss events, this study is proposing that people discount the good luck embedded in others' decision making and their subsequent outcomes. We are proposing that neutral third parties will discount the fact that near-misses are derived in part from good luck, and will categorize these

near-miss events just as they would true misses.

Hypothesis 2a: Managers whose decisions result in a near-miss will have their decision making evaluated more favorably than managers whose decisions result in a hit and less favorably than managers whose decisions result in a miss.

Hypothesis 2b: Managers whose decisions result in a near-miss will be judged more competent, intelligent, to have more leadership ability, and to be more promotable than managers whose decisions result in a hit and judged less competent, less intelligent, to have less leadership ability, and to be less promotable than managers whose decisions result in a miss.

Hypothesis 3: Managers whose decisions result in a near-miss will be judged closer to those whose decisions ended in a miss than to those whose decisions ended in a hit.

Attenuating the Biases

Some studies have shown the outcome bias can be attenuated when people have high accuracy motives, rather than low accuracy motives (Allison, Beggan, McDonald, & Rettew, 1995). More recently, Agrawal and Maheswaran (2005) have shown that people with "accuracy" goals, who wish to be accurate in their judgments, show less outcome bias than people with "defensive" goals who want to defend an impression that is consistent with the outcome. The rationale here is that a desire for accuracy induces an elaboration of objective information, that is, a more balanced assessment of any situation. The more detailed the decision making, the more objective elaboration required, and hence the more balanced the ensuing decisions. This is thought to minimize the effects of heuristics and biases. Thus we propose that when people have to make more detailed judgments about specific decisions, these judgments

will show less bias than those of general decision making, competence, intelligence, leadership, or promotability.

Hypothesis 4: Favorability evaluations of managers' specific decisions will show less outcome and near-miss biases than favorability evaluations of managers' general decision making, competence, intelligence, leadership, and promotability.

Method: Materials, Participants, and Procedure

This research builds on historical mission-case data currently available and being developed at Goddard Space Flight Center (GSFC) to investigate what NASA has learned from past mission successes, near-misses, and failures.[1] A case scenario was created loosely based on development details from the two NASA missions: TIMED and WIRE. TIMED (Thermosphere Ionosphere Mesosphere Energetics and Dynamics Project) was an ambitious project when conceived in 1990 involving multiple spacecraft and nine instruments. When finally launched in 2001, it was one spacecraft with four instruments. Development of the project shifted between GSFC and the Applied Physics Lab (APL), and the final four years of the project were a joint effort of both centers (1990-1993 – GSFC project, 1994-1996 – APL project, 1997-2001 – APL/GSFC project). These management shifts resulted in confusion over reporting relationships and responsibility. Also, in several documented cases, GSFC managers felt that APL processes were not up to GSFC standards. Immediately after launch, the spacecraft experienced four guidance and control anomalies, but the operations team was able to quickly overcome the problems, and the mission met its level one science requirements and was therefore categorized a success. WIRE (Wide-Field Infrared Explorer) was a project within the Small Explorer (SMEX)

11

[1] Full case studies are available at the website of the Office of Mission Success at: www.missionsuccess.gsfc.nasa.gov.

program office. The SMEX program has been considered a very successful program with the exception of the WIRE mission failure. In the case of WIRE, GSFC was responsible for the overall mission, but the Jet Propulsion Laboratory (JPL) was responsible for the instrument development. JPL contracted with a third party to develop the instrument electronics (including pyro control electronics). On-going "sibling rivalries" between GSFC and JPL and confusion over reporting relationships led to poor communications and responsibility conflicts for the project. Also, fairly late in the process, there was turnover in the instrument manager position at JPL. WIRE failed soon after launch in March 1999 when the cover on the instrument ejected prematurely and destabilized the spacecraft. The spacecraft was recovered and completed some investigations but did not achieve its primary science objectives. It was thus considered a failure.

In our case scenario, three mission versions were created (miss, near-miss, and hit). In all three cases, the development problems were identical, and Chris, a project manager makes the same decisions about how to interact across the different NASA centers, whether or not to skip a peer review and whether or not to delay the mission to investigate an unlikely problem but one that had the potential to be fatal for the mission. Appendix 1 shows the actual text of the scenarios.[2] In all three cases, because of turnover and tight schedules, Chris allowed the project to miss a peer review of the electronics of the instrument and decided not to investigate a last minute, potentially catastrophic (albeit low probability) design problem. The cases differed only in project outcome. In the "miss" version, the reader is told that the mission was a complete success sending back useful data (i.e., there is no error shortly after launch). In the "near-miss version", a problem occurs shortly after launch, and only because of the spacecraft's alignment

12

[2] In the case write-up given to students, the mission was originally named WIRE after the actual NASA mission. This was not a problem because no students were familiar with the actual mission. The name was changed to a

to the sun, the problem is NOT catastrophic to the mission. In the "hit" version, the same problem occurs shortly after launch and only because of the spacecraft's alignment to the sun the problem IS catastrophic.

After reading the case scenario, participants rated Chris' decision making ability, competence, intelligence, leadership, and whether Chris should be promoted to a larger project and or whether Chris should be fired. They also rated two of Chris' specific decisions: the decision to skip the peer review and the decision not to investigate a last minute, low probability, venting problem. All ratings were done on a 7 point Likert scale, with 7 being most positive and 1 being least positive. Participants were also asked to supply demographic information (age, gender) as well as project management and general people management experience.

Data were collected from three groups of participants: NASA managers (N=24), MBA students at a large eastern university (N=95), and undergraduate business students at that same university (N=83). NASA managers filled out the exercise as part of a training course on project management. The students filled out the exercise as part of their classroom activities, during a lesson on general decision making.

Analysis and Results

Control variables

Bivariate correlation showed that, as expected, the dependent variables were significantly correlated with each other (see Table 1). The analysis also showed that age was correlated to three dependent variables: intelligence ($p<.01$), the decision not to investigate the possible vent issue ($p<.05$), and whether or not Chris should be fired ($p<.01$), and managerial experience was correlated to whether or not Chris should be fired ($p<.05$). Age and managerial experience

fictitious name in the version provide to NASA managers.

differed significantly across the three samples (age: NASA managers: mean = 43.1, standard deviation = 7.77; MBA mean = 29.4, standard deviation = 3.31; undergraduate mean = 19.4, standard deviation = 0.96; $F(2, 135) = 412$; $p<.001$; managed people: NASA managers = 84%; MBA = 72%; Undergraduate = 32%). Thus to conserve degrees of freedom, "sample" was used as a covariate in the analyses to control for any response differences due to age and managerial experience.

A MANOVA was performed since the dependent variables (decision ability, competence, intelligence, etc.) are correlated (Devore, 1987). Condition (miss, near-miss, hit) was the independent variable and sample was entered as a covariate.

Sample had a significant impact on the data. Univariate F statistics were marginally significant for competence ($F(2, 198) = 2.8$, $p=.09$) and intelligence ($F(2, 198) = 3.1$, $p=.08$) and significant for leadership ability ($F(2, 198)= 7.9$, $p<.01$) and whether or not to fire Chris ($F(2, 198) = 10.6$, $p=.001$). Post-hoc contrasts, using Tukey HSD, showed that these univariate differences were driven by NASA managers having rated Chris as significantly more competent (NASA mean = 4.8, standard deviation = 1.17; MBA mean = 4.3, standard deviation = 1.16; UG mean = 4.2., standard deviation = 1.17) and more intelligent (NASA mean = 5.13, standard deviation = 1.22; MBA mean= 4.1, standard deviation = 1.22; UG mean = 4.3, standard deviation = 1.22) than the other two groups. The MBA managers rated Chris to have significantly less leadership ability (NASA mean = 5.3, standard deviation = 1.41; MBA mean = 4.7, standard deviation = 1.40; UG mean = 5.2, standard deviation = 1.40) than the other two groups, and the NASA managers were significantly less likely to think firing Chris would be appropriate (NASA mean = 2.6, standard deviation = 1.5; MBA mean = 4.2, standard deviation = 1.49; UG mean = 4.3, standard deviation = 1.49). Thus it appears that, in general, the NASA

managers tended to rate Chris a bit more favorably than the other two groups.

Very importantly, however, there were no significant interaction effects between sample and condition, meaning that the pattern of responses was not different across the different samples, i.e., the effect of condition on ratings of Chris and Chris' decisions were the same across all the samples examined. Thus for hypothesis testing, all participants were used, and while sample was used as a covariate in the analysis, we report only results from all participants together.

Hypothesis testing

The MANOVA with condition as an independent variable and sample as a covariate had a significant multivariate F for condition (test: Wilks' Lambda(16, 382)= 2.64, $p<.001$) meaning that the ratings of the project manager across the three cases (miss, near-miss, hit) were different. Univariate F statistics showed significant differences for decision ability ($F(2, 198)= 5.3$, $p<.01$), competence ($F(2, 198)= 7.7$, $p=.001$), leadership ability ($F(2, 198)= 4.2$, $p<.05$), and the decision to promote the manager to a larger project ($F(2, 198)= 15.2$, $p<.001$). Univariate F statistics showed significant differences for participant's assessments of Chris' specific decisions: peer review decision ($F(2, 198)= 3.5$, $p<.05$), and the decision to ignore potential vent issue ($F(2, 198)= 6.7$, $p=.001$). No significant differences were seen in an assessment of the project manager's intelligence or in whether or not to fire the manager following the outcome of the project across conditions. Post-hoc contrasts, using Tukey HSD, were used to test hypotheses as to which conditions were significant from each other. Table 2 summarizes these statistical comparisons for the dependent variables by condition and Figure 1 provides a graphical representation of the mean responses by condition.

Hypothesis 1a proposed managers whose decisions result in a miss will have their decision making evaluated in a significantly more favorable light than managers whose decisions result in a hit. This was supported. Managers whose decisions result in a miss had a decision ability average of 4.7 (standard deviation = 1.57), whereas managers whose decisions result in a hit had a decision ability average of 3.8 (standard deviation = 1.75) (Tukey HSD, p=.006).

Hypothesis 1b proposed that managers whose decisions result in a miss will be judged more competent, intelligent, to have more leadership ability, and to be more promotable than managers whose decisions result in a hit. This hypothesis was mostly supported; it was supported for competence (miss mean = 4.7, standard deviation = 1.27; hit mean = 3.9, standard deviation = 1.28, p=.001), leadership ability (miss mean= 5.4, standard deviation = 1.18, hit mean=4.6, standard deviation = 1.62, p=.01), and promotability (miss mean = 3.9, standard deviation = 1.58; hit mean = 2.7, standard deviation = 1.48, p<.001), but not for general intelligence.

Hypothesis 2a proposed that managers whose decisions result in a near-miss will have their decision making evaluated more favorably than managers whose decisions result in a hit and less favorably than managers whose decisions result in a miss. This was partially supported. The hit condition (mean = 3.8, standard deviation = 1.75) was significantly different from the near-miss condition (mean = 4.4, standard deviation = 1.61; p=.05), but the near-miss and miss conditions were not significantly different from each other.

Hypothesis 2b proposed that managers whose decisions result in a near-miss will be judged more competent, intelligent, to have more leadership ability, and to be more promotable than managers whose decisions result in a hit and judged less competent, intelligent, to have less leadership ability, and to be less promotable than managers whose decisions result in a miss.

This was partially supported. Just as above, for competence, the hit condition was significantly different from the near-miss condition (mean = 4.4, standard deviation = 1.19, p=.007), but the near-miss condition was not significantly different from the miss condition. Again for promotability, the hit condition was significantly different from the near-miss condition (mean = 3.5, standard deviation = 1.53; p=.001), but the near-miss condition and miss condition were not significantly different from each other.

Hypothesis 3 proposed that managers whose decisions result in a near-miss will be evaluated closer to those whose decisions ended in a miss than to those whose decisions ended in a hit. The pattern of responses for Hypotheses 2a and 2b confirms Hypothesis 3. For decision ability, competence, and promotability, the near-miss condition was significantly different from the miss condition but not significantly different from the hit condition. The data show that managers whose decisions result in a near-miss are evaluated very similarly to those whose decisions end in a miss, but significantly differently from those whose decisions result in a hit.

Hypothesis 4 proposed that favorability evaluations of managers' specific decisions will show less outcome and near-miss biases than favorability evaluations of managers' general decision making, competence, intelligence, leadership, and promotability. As noted above, the univariate F statistics showed significant differences for participant's assessments of Chris' specific decisions: peer review decision ($F(2, 198)= 3.5$, $p<.05$), decision to ignore potential vent issue ($F(2, 198)= 6.7$, $p=.001$), suggesting little support for this hypothesis. Post hoc comparisons showed that for the peer review decision, the near-miss (mean = 2.9, standard deviation = 1.37) was significantly different from the hit condition (mean = 2.4, standard deviation = 1.29; p=0.04), but was not significantly different from the miss condition (mean = 2.9, standard deviation = 1.27). For the decision to ignore the potential vent issue and launch as

scheduled, the miss condition (mean = 3.6, standard deviation = 1.67) was significantly different from the hit condition (mean = 2.7, standard deviation = 1.51; p=.001), and the near-miss condition (mean=3.2, standard deviation = 1.73) was marginally different from the hit condition (p=.10) but not significantly different from the miss condition. Thus, overall, it appears that the pattern of data for the evaluations of the specific decisions mirrors that found in the more general decisions, thus hypothesis 4 is not supported.

Discussion

Participants rated project managers whose decisions ended in a miss (organizational success) in a significantly more positive light than they rated project managers whose decision ended in a hit (organizational failure). Supporting H1, participants rated project managers with miss outcomes as having significantly higher decision ability than project managers with hit outcomes. Supporting H2, this outcome bias spilled over to other characteristics of the project managers, such that those managers with miss outcomes were rated as significantly more competent, having more leadership ability, and being more deserving of a promotion than project managers with hit outcomes. The lack of differences for intelligence may stem from the fact that this construct is too general. Outcome valence may bias people's perception of decision quality and other domain-related (here, project management- related) attributes, such as competence, leadership, and promotability. Whereas, context-free characteristics like general intelligence may be more immune to outcome bias.

Aside from the general outcome bias, the results for how near-miss events were categorized shows evaluations were also influenced by a near-miss bias. Partial support for H2a and H2b and the general support of H3 showed that project managers whose decisions resulted in

a near-miss were judged in between project managers with miss outcomes and those with hit outcomes, yet were categorized closer to project managers with miss outcomes than those with hit outcomes. In fact in this data, there were few significant differences in the categorization of near-misses and misses, and quite a few significant differences in the categorization of near-misses and hits. This skew in the categorization of near-miss events is quite important, explaining why we claim "bias" in the processing of near-miss events. If near-miss events were processed objectively, then they should be categorized as either hits or misses, depending on how close they were to being actual hits or misses, that is—depending on the extent to which good luck prevented a near-miss from being a hit. Over a diverse group of participants, we would have expected some to categorize near-misses as misses and some to categorize near-misses as hits. The fact that near-miss events were systematically being categorized as misses (and being distinct from hits), suggests people are discounting the luck component of a decision's outcome. Again this is consistent with prior work which showed that people discounted the luck involved in their own near-miss events (that they were not fatalities) (Tinsley & Dillon, 2005). Here we show that objective, neutral observers discounted the luck involved in others' decision outcomes. That is, participants here discounted others' good luck in achieving a near-miss rather than a hit.

This is somewhat remarkable since prior work on attribution errors, suggests that individuals will discount their own good luck, making internal attributions for their own good outcomes (the so-called, "self-serving attributional bias", Miller & Ross, 1975) but be more resistant to discounting others' good luck, as they are able to discern external reasons for others' good outcomes. Lau and Russell (1980) for example have shown that the self-serving attributional bias (the tendency to make internal attributions for one's own success) extends to close friends and other groups to which one is allied, but does not generally extend beyond this

set of in-group members. Our participants (particularly the students) were in no sense in-group members with the project managers they were evaluating, thus there should be no impulse to make internal attributions for positive outcomes and discount the component of good luck in these outcomes. Yet, our participants did still discount the component of good luck in the project managers' near-miss outcomes, believing that these near-miss outcomes implied the project managers had good decision ability, competence, leadership ability and promotability.

Additionally, while our sample size of NASA managers was too small to statistically examine their responses exclusively, the trend (as can be seen in Figure 1) is for NASA managers to actually evaluate Chris in the near-miss case as higher than Chris in the miss case. This certainly may be consistent with norm and prospect theory in that in the near-miss case, participants are comparing it to a success and mentally anchoring on that outcome, while in the miss case, no comparisons are invoked. Additional data collection exercises are already scheduled with NASA, so we will have the opportunity to examine this behavior further in future research.

Attenuating the Bias

Future research should look at factors that might attenuate this near-miss bias. We reasoned that an elaboration of objective information might produce a more balanced assessment of the near-miss events. The more detailed the decision making, the more objective elaboration required, and hence we reasoned that participants judgments of project managers' specific decision might show less outcome bias and less near-miss bias than their evaluations of project managers' general decision making ability (H4). Unfortunately, our data did not support this. It is possible that our within subjects design led to this lack of support. Once participants had

evaluated a project managers general ability as either high or low (depending on whether the project manager had a miss, near-miss, or hit outcome), cognitive consistency impelled them to judge the specific decision in a similar fashion. A between subjects design, where some participants judge only the specific decisions made, might show that evaluation of more specific decisions shows less bias than evaluation of general decision making.

Another condition that might attenuate the near-miss bias would be a prime to make salient the good luck component of any decision outcome. One way to make luck salient might be to ask participants to imagine alternative outcomes to the near-miss event, and how and why the near-miss event is not a hit. Mutable (changeable) factors that explain why a near-miss event is not a hit should prompt people to more readily imagine the near-miss event as a hit, as norm theory dictates that mutable factors between an event and it's normative alternative increase the perceived "closeness" of the event and its normative alternative (Kahneman & Miller, 1986). "Mutable" here does not mean that it is within the power of the project manager to make the changes that turn the event (near-miss) into the normative alternative (hit), only that it is imaginable that a factor could change. In our scenario, the mutable factor is the alignment of the sun with the spacecraft. In the near-miss condition, the alignment was said to mitigate the damage done by the venting problem, in the hit condition the alignment was said to further the damage done by the venting problem. The sun's alignment is mutable if participants could imagine that the sun might be aligned differently so as to cause a hit rather than a near-miss.

Implications for Organizational Learning and Organizational Culture

The outcome bias and near-miss bias will clearly impact organizational learning. When NASA experiences a failure or hit, a formal investigation board is convened to identify the

factors that contributed to the outcome. Yet, as we suspect happens in most organizations, near-misses do not command the same type of attention. For example, within the space shuttle program, much of the prior debris problems could have caused a similar catastrophic failure as it did on the Columbia mission except that on previous missions, everyone was lucky that a large piece did not hit a highly sensitive portion of the orbiter. Thus, with hindsight, we might suggest that the Columbia accident was partly caused by poor decision-making and partly by bad luck. The shuttle experienced bad luck in that the piece of foam debris struck the leading edge of the wing, a very sensitive portion of the orbiter. However, poor decision-making by project managers allowed similar pieces of foam to become detached at least thirty times on previous missions and only good luck prevented a large enough piece of foam from striking a highly sensitive area on any of these previous shuttle missions (CAIB, 2003). Without the obvious technical failure, everyone interpreted the near-miss events as successful missions, and this acceptance of foam debris was adopted as a normal occurrence by the shuttle program managers. Hence, what was at one point a cause for concern (the debris) because a normal occurrence. Deviance became normalized (Vaughn, 1996). Hence the near-miss bias explains, in part, how a culture may come to embrace more risk, or deviance, and how it fails to learn what steps to take to mitigate that risk.

Another mechanism by which the near-miss bias may become instantiated in a culture and lead to a culture of risk taking, involves the risk propensities of the managers who get promoted. Recall that in this study, project managers whose decisions resulted in a near-miss were judged as significantly more competent and deserving of a promotion than project managers whose decisions result in a hit. This difference occurred even though it was quite clear in both conditions that the spacecraft did experience a venting problem and it was clear that

it was only because of the sun's alignment that the outcome was either a near-miss or a hit. This suggests that even when a problem occurs down the road that is reasonably linked to prior managerial decisions (in both cases Chris makes the decision not to investigate the venting issue), that manager is not held accountable for any faulty decision making if the project outcome is unharmed. Thus both that manager and the organization as a whole miss a learning opportunity for collecting data on how to mitigate future risks.

Moreover, our data showed that managers whose decisions end in misses or near-misses were judged significantly more competent, having more leadership and decision making ability, and being more deserving of promotions than those whose decision ended in a hit. Hence, managers who experience near-misses are more likely to move up the corporate ladder than managers whose decisions end in unsuccessful outcomes. Prior work documents that people who had experienced a near-miss made significantly more risky decisions in the future, than those who had not experienced a near-miss (Tinsley & Dillon, 2005). Assuming that those who experience near-misses make riskier subsequent decisions and that those whose decisions end in near-misses move up the corporate ladder, than those who are promoted should, in general, embody more risk that those who are not promoted (experienced bad luck and had a hit). Assuming the level of risk reflected in an organization's decisions is more heavily weighted by the risk tolerance of those at high rather than low levels of the organizations, then near-miss events will bias organizations towards riskier decisions over time as risk tolerant managers advance. The near-miss bias thus becomes institutionalized and embodied within an organization.

Conclusion

The evaluation of a manager's decision making reflects both an outcome bias and a near-miss bias. Managers whose decisions ended in a failure were perceived as having significantly less decision making ability, being significantly less competent, showing significantly less leadership ability and being significantly less deserving of promotions than managers who made the same decisions but whose project outcomes were either a success or a near-miss. Strikingly, there were no significant differences between the miss (success) and near-miss condition, rather near-misses appear to be systematically categorized as miss events rather than hit (failure) events. The differences between the near-miss and hit conditions mean that even when a problem occurs down the road that is readily linked to prior managerial decisions, if the project outcome is not harmed because of luck then managers experiencing the good luck are not held accountable for any faulty decision making. This severely hampers the learning potential from near-miss events for both the managers and the organization, and it suggests a mechanism by which the near-miss bias becomes instantiated in an organizational culture.

REFERENCES

Agrawal, N. & Maheswaran, D. 2005. Motivated reasoning in outcome-bias effects. *Journal of Consumer Research, 31*: 4-14.

Allison, S.T., Beggan, J.K., McDonald, R.A., & Rettew, M.L. 1995. The belief in majority determination of group decision outcomes. Basic and Applied Social Psychology, 16 (April): 367-382.

Allison, S.T., Mackie, D.M. & Messick, D.M. 1996. Outcome biases in social perception: Implications for dispositional inference, attitude change, and social behavior. In M.P. Zanna (Ed.), *Advances in experimental social psychology, 28:* 53-93. New York: Academic Press

Baron, J.& Hershey, J.C. 1988. Outcome bias in decision evaluation. *Journal of Personality and Social Psychology, 54:* 569-579.

Columbia Accident Investigation Board Report. 2003. *Final Report, Volume 1.* Washington, DC: NASA and the Government Printing Office.

Dawes, R.M. 1988. *Rational choice in an uncertain world.* New York: Harcourt, Brace, Jovanovich.

Devore, J.L. 1987. *Probability and Statistics for Engineering and the Sciences*, 2nd edition, Wadswach, Inc., California.

Dillon, R.L. and Tinsley, C.H. 2005. Interpreting near-miss events. *Engineering Management Journal,* April 2005.

Edwards, W., Kiss, I., Majone, G, & Toda, M. 1984. What constitutes a 'good decision'? Panel discussions with introduction and postscript by Charles Vlek, *Acta Psychologica, 56* (August), 5-27.

Festinger, L. 1957. A theory of cognitive dissonance. Evanston, IL: Row, Peterson.

Fischoff, B. 1982. Debiasing. In D. Kahneman, P.Slovic, and A. Tversky (Eds.), Judgment under uncertainty: Heuristics and Biases. New York: Cambridge University Press: 422-444.

Fischoff, B. & Beyth, R. 1975. I knew it would happen—Remembered probabilities of once future things. *Organizational Behavior and Human Performance, 13:* 1-16.

Galinsky, A. D., G. Moskowitz, and I. Skurnik. 2000. "Counterfactuals as self-generated primes: The effects of prior counterfactual activation on person perception judgments." *Social Cognition.* 18: 252-280

Jensen, M.C. and Meckling, W.H. 1976. Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure, *Journal of Financial Economics*, 3: XX.

Kahneman, D. & Miller, D.T. 1986. Norm theory: Comparing reality to its alternatives. Psychological Review, 93: 136-153.

Kahneman, D. & Tversky, A. 1979. Prospect theory: An analysis of decision under risk. *Econometrica,* 47: 263-291.

MacCormack, A. (2004). Management Lessons from Mars: Go ahead and raise the bar. Just don't make the same mistakes NASA did, *Harvard Business Review*, May 2004, pp.18-19.

Miller, D.T. & MacFarland, C. 1986. Counterfactual thinking and victim compensation : A test of norm theory. *Personality and Social Psychology Bulletin, 12*: 513-519.

Miller, D.T. & Ross, M. 1975. Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin, 82*: 213-225.

Lau, R.R. & Russell, D. 1980. Attributions in the sports pages. *Journal of Personality and Social Psychology, 39:* 29-38.

Tinsley, C.H. & Dillon, R.L. 2005. Whew That Was Close: How Near-Miss Events Bias Subsequent Decision Making Under Risk. Working paper, Georgetown University Center

for Business and Public Policy.

Vaughan, D. 1996. *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*. Chicago: University of Chicago Press.