# COMMUNICATING DATA-DRIVEN RISK INFORMATION TO PILOTS

Nicoletta Fala
Purdue University
West Lafayette, IN

Karen Marais
Purdue University
West Lafayette, IN

General Aviation safety is a pressing concern. In this research, we consider the factor that appears most often in accidents: the pilot. Newly-licensed pilots can fly without their instructor, potentially as the only or most experienced pilot in the aircraft. Commercial debrief products use technology in the flight deck to collect data and provide post-flight visualizations for performance reviews, but do not discuss flight safety. To manage risk, though, pilots need to perceive the risk associated with a situation before deciding whether they are willing to accept it. Safety-driven post-flight feedback may help address performance. However, it is not clear whether and how the way we present feedback affects how pilots perceive risk, or what the best way is. We designed and disseminated a survey to evaluate the communication factors that affect pilots' risk perception. In this paper, we evaluate whether different representation methods affect relative risk perception among pilots.

General Aviation (GA) consists of all civilian aircraft operations other than commercial air transport operations and it covers a range of activities, both commercial (business aviation) and non-commercial (recreational and flight training operations). In 2017, GA in the United States was responsible for 990 non-commercial fixed-wing accidents (AOPA, 2018a). Non-commercial GA, in particular, has been contributing disproportionately to the aviation accident rate, with an accident rate of 5.57 accidents per 100,000 flight hours, significantly higher than the rate of 2.33 accidents per 100,000 flight hours in commercial fixed-wing GA. Similarly, 20.3% of non-commercial fixed-wing accidents are fatal; almost double the 10.4% of commercial fixed-wing operations. Most GA accidents (~74%) are attributed to pilot-related causes—they occur because of the pilot's action or inaction.

Continuing to provide pilots with feedback on their flying even after they finish their training and are no longer flying with an instructor (and potentially flying as the sole pilot or the most experienced pilot in the aircraft) can improve GA safety. Rantz et al. (2009) evaluated how feedback and praise can be used to increase the extent to which pilots use checklists accurately, with some participants in their study showing abrupt improvements in performance after intervention. Commercial products that leverage the addition of technology in the cockpits of small aircraft to collect flight data and present pilots with a visualization of their flights, like CloudAhoy and CirrusReports, are becoming more prevalent. However, these products do not discuss risk or flight safety. O'Hare's Aeronautical Risk Judgment Questionnaire (ARJQ) suggests that pilots display low levels of risk and hazard awareness, and an optimistic self-appraisal of their abilities (O'Hare, 1990). If pilots do not identify risk that can be mitigated in

their flying, we cannot expect that they will improve. To manage risk, pilots need to perceive the risk associated with a situation or hazard and decide whether they are willing to accept that level of risk in each situation (Hunter, 2002). Safety-driven post-flight feedback may help facilitate risk management in subsequent flights, by alerting pilots to potentially hazardous situations. However, no one has considered what the best way is to present risk-related feedback in the pilot population, and we do not know whether presentation format affects how they perceive risk.

Using flight data to proactively improve GA safety requires that we are able to (1) identify behaviors that may put the safe outcome of a flight at risk, (2) detect those behaviors in the available flight data, and (3) inform the pilot in a way that helps them improve in their future flights. We use a state-based representation of historical aviation accidents to define a list of undesirable events or behaviors that we need to communicate to the pilots, in the form of states and triggers. Each flight consists of states, which can be nominal or hazardous, and trigger events (Rao, 2016). A state is a period of time during which the system, consisting of the aircraft and the pilot, exhibits a particular behavior, and a trigger is an event that causes the system to transition between two states. We use flight data to retrospectively detect these states and triggers, upon completion of the flight, by mapping parameters or combinations of parameters that can be tracked in the flight data to the hazardous states and triggers defined. We then present any detected hazardous states to pilots in the form of post-flight debrief feedback, with the goal of using the information to improve safety on subsequent flights. To evaluate the effectiveness of feedback in different representation formats, we used an anonymous web-based survey where a sample of pilots self-debriefed flights with safety information presented in different ways, and assessed the risk of the flight in each case. We also asked the pilots how likely they are to make changes to their flying as a result of the information they reviewed, to evaluate feedback effectiveness in terms of motivation to change unsafe behaviors. We demonstrated this approach using the hazardous states that are specific to the takeoff phase of flight.

## Cognitive Biases in Risk Perception Among Pilots

We hypothesize that pilots will perceive the risk of their flight depending on how information is presented to them. While research in the medical, education, and economics fields have established guidelines that designers can use when communicating risk to the general population, the pilot sub-population is understudied, so we do not know which cognitive biases affect their understanding. We consider three factors that may impact their risk perception: framing language, representation method, and parameter type. Framing language corresponds to whether we discuss risk in terms of safety-centric language or risk-centric language. For example, we might refer to a safe flight as being either ‘very safe’ or ‘not risky’. Representation methods refer to how we present data: graphically or numerically/textually. For example, we can communicate how much runway distance was remaining at takeoff numerically (2,500 ft), or graphically on the airport diagram. Lastly, parameter type refers to how we frame the same metric. For example, there are two ways to measure deviation from the runway centerline: the distance between the aircraft’s longitudinal axis and the runway centerline, or the distance from the aircraft’s longitudinal axis to the edge of the runway. The former represents how close the pilot was to the ideal condition (with the aircraft’s longitudinal axis aligned with the centerline); the latter measures how close the pilot was to being involved in an incident (runway excursion or

collission with an object). While both versions represent the same thing mathematically, parameter type can affect how pilots perceive their own flight performance in terms of risk.

To investigate how these three factors impact safety-driven feedback effectiveness, we created a $2^3$ full-factorial design experiment. Table 1 shows the eight resulting treatment combinations. Using de-identified flight data from a Garmin G1000 display and the CloudAhoy flight visualization software, we designed interactive prototype debrief screens that include safety-driven feedback. Figure 1 shows how we altered the commercial screens to add risk information in two popup windows; (a, top right) provides the pilot with a list of behaviors that could potentially appear in a flight, and (b, center) displays parameters that characterize the selected behavior. The survey is available at www.nicolettafala.com/survey and an example of an interactive prototype is available at www.nicolettafala.com/debriefexample. We repeated this process for three different flights, and survey respondents could respond to as many of the three flights as they wished.

Table 1.
*The $2^3$ full-factorial design evaluates main and interaction effects among the three factors.*

| Treatment Group | Framing Language | Representation Method | Parameter Type | Responses [Flight A] |
|---|---|---|---|---|
| 1 | safety-centric | graphical | performance | 35 |
| 2 | risk-centric | graphical | performance | 31 |
| 3 | safety-centric | numerical | performance | 44 |
| 4 | risk-centric | numerical | performance | 33 |
| 5 | safety-centric | graphical | safety | 23 |
| 6 | risk-centric | graphical | safety | 34 |
| 7 | safety-centric | numerical | safety | 33 |
| 8 | risk-centric | numerical | safety | 35 |

To evaluate feedback effectiveness, we asked respondents six questions: (Q1) Given the information presented to you, how safe would you say this takeoff was? [5-point Likert scale] (Q2) In this takeoff, which of the following would concern you, if any? [Centerline deviation, Rotation airspeed, Engine RPM, Takeoff distance, Wind] (Q3) Optional Comments. (Q4) What changes (up to 5) do you think you could make to an upcoming flight after the information presented here, if any? [Freeform text, up to 5 changes] (Q5) How likely are you to make each of these changes to an upcoming flight? [5-point Likert scale for each change] (Q6) How important do you think each of these changes is to improving safety on takeoff? [5-point Likert scale for each change] To evaluate the impact of a risk-centric framing language, we reworded these questions, replacing safe with risky (How risky would you say this takeoff was).

We measure feedback effectiveness in two ways: (1) did the pilot understand how safe or unsafe their flight was based on their responses to Q1 and Q2, and (2) how motivated are they to change their behaviors to mitigate the risk in their flying activities based on responses to Q4, Q5, and Q6? Q1 captures how different treatment groups introduce cognitive biases in pilots and is subjective—we expect to see differences in the distributions of responses among the different treatment combinations. Q2 can identify whether pilots are perceiving risk in the correct

categories more objectively; that is, if we deem a flight to be unsafe due to high crosswinds, did the pilots identify high crosswinds as an issue?

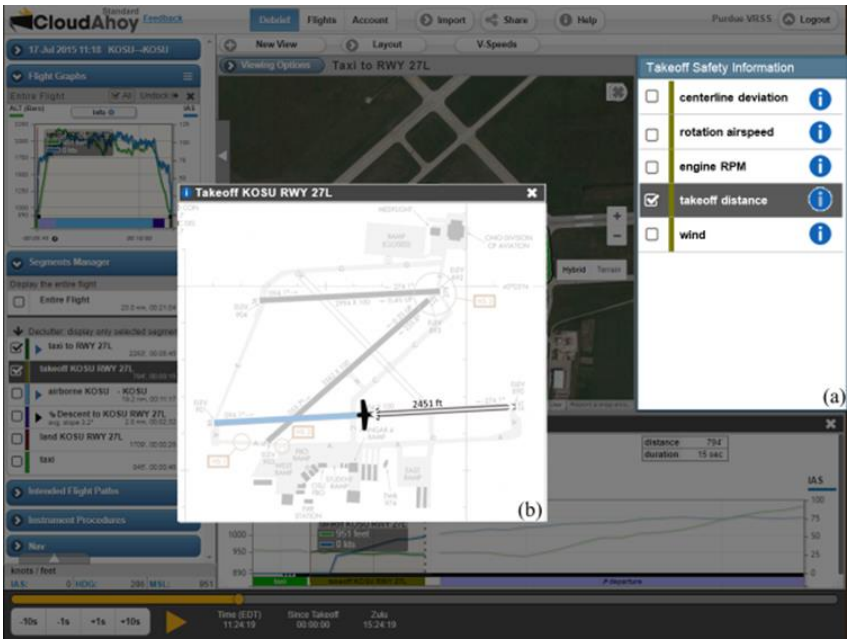


*Figure 1*. We supplemented the visualization of the flight data from CloudAhoy with information on the safety of the flight on five parameters.

## Survey Results and Analysis

We used aviation mailing lists and groups to recruit participants and encouraged snowball sampling to maximize our responses. Approximately 70% of our respondents provided us with demographic information. Our sample consisted of 71% male and 26% female pilots, 76% of whom have completed at least a 4-year degree. Private pilots and commercial pilots made up the majority of the sample, at 49% and 30% respectively, with 58% of the pilots being instrument-rated. Most respondents (64%) fly primarily aircraft with steam gauges, fly at least weekly (59%) and have never used commercial debrief or flight visualization products like CloudAhoy (88%). We ended up with 268 responses for the first flight scenario, 195 for the second flight, and 189 for the third flight. Since the first flight in the survey got the maximum number of responses, we focus our initial data analysis on that flight before comparing results with the other two flights.

Figure 2 depicts the raw responses to Q1 that correspond to each group—the darker the marker, the higher the frequency of that particular response. The average response (marked in orange) in each treatment group tends to oscillate around the neutral response of 3 out of 5 on the risk Likert scale, but the spreads are also different, with the mean of the eighth group appearing lower than the means of the other treatment groups. We processed the data for the first flight to identify the response to Q1 and the treatment group it belongs to for each respondent and used ANOVA to test for the difference in treatment groups. The probability of the response means being equal for different treatment groups is < 2%. We also ran the Tukey procedure to identify which treatment groups had significantly different responses.

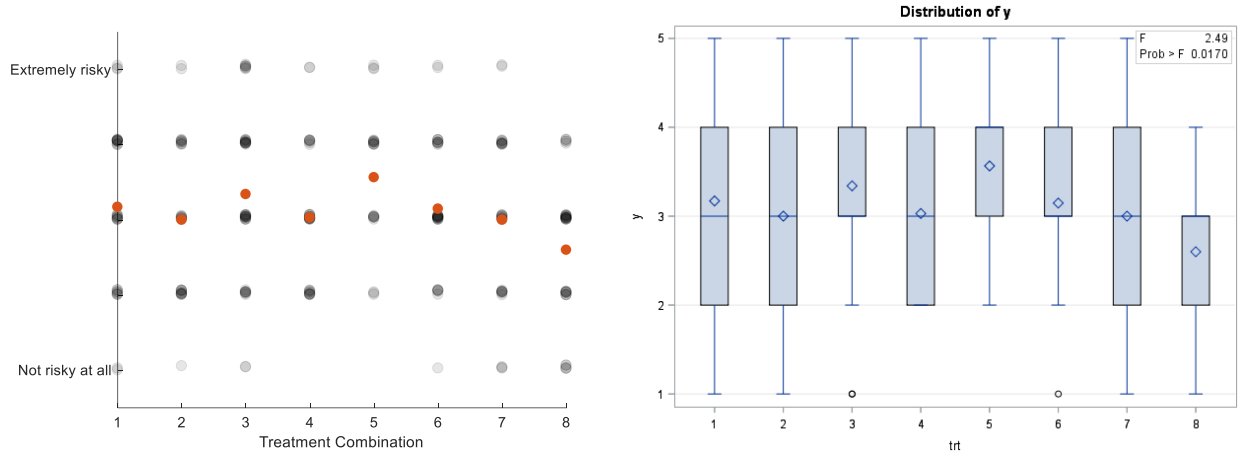


Figure 2. *The responses for the first flight show means and spreads that visually vary slightly among the eight treatment combinations. Treatment group 8 biased the respondents towards a less risk-averse response, and treatment groups 3 and 5 show a more risk-averse bias.*

Our ANOVA test indicates that there is a difference between the eight groups, with a p-value of 0.0170. Table 2 summarizes the ANOVA results. The box plot in Figure 2 suggests that the means for treatment groups 3 and 5 are higher than the mean for treatment group 8. Tukey's HSD test, used in conjunction with the ANOVA results, compared all possible pairs of means and identified which ones are significantly different from each other. The test identified that some treatment combinations—3 and 8 and 5 and 8—are different at the 0.05 significance level. Table 3 shows the difference between the means of these groups. Treatment combinations 3 and 5 have the safety-centric framing language in common, whereas treatment combination 8 is framed in terms of risk. This discrepancy suggests that asking a pilot how safe their flight was vs. how risky their flight was can potentially make the pilot more risk-averse.

Table 2.
*We reject $H_0$ based on the ANOVA results. The means differ between the different groups.*

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 7 | 17.3332430 | 2.4761776 | 2.49 | 0.0170 |
| Error | 260 | 258.1443690 | 0.9928630 | | |
| Corrected Total | 267 | 275.4776119 | | | |

Table 3.
*Tukey's studentized range (HSD) test for the response variable at $\alpha = 0.05$. The table only shows those treatment groups that are considered statistically different.*

| Treatment Group Comparison | Difference Between Means | Simultaneous 95% Confidence Limits | |
|---|---|---|---|
| 5 – 8 | 0.9652 | 0.1479 | 1.7825 |
| 3 – 8 | 0.7409 | 0.0513 | 1.4305 |

We ran a second ANOVA test to evaluate whether pilots responded differently to messages framed in a risk-centric language and safety-centric language. The probability of the

two means being equal is approximately 1%, with the safety-centric language moving the location of the mean towards a more risk-averse response.

## Conclusion and Future Work

Our initial results indicate that pilots are subject to certain cognitive biases that will impact the way they perceive their flight risk. The limitation to this work is that the results are based on one flight that is repeated for all participants—if the level of risk in the flight or the specific states in the flight affect risk perception, our results may be affected. The next step for this research, therefore, is to analyze different flights (the second and third flight in our survey) to evaluate whether the conclusions are flight-specific or valid across the board. We will then investigate the second part of our definition of "feedback effectiveness"—does the way we present safety information impact how motivated pilots are to change their behaviors? Our response variable metrics for motivation are (a) the number of changes suggested, (b) the average willingness to change among the suggested changes, and (c) the maximum willingness to change among the suggested changes.

## Acknowledgments

We thank the Purdue Statistical Consulting Service for their help with the data analysis, and the people and organizations that were significantly helpful in our snowball sampling (the Ninety-Nines and the Cardinal Flyers in particular).

## References

AOPA Air Safety Institute (2018a). GA Accident Scorecard. Retrieved from https://www.aopa.org/-/media/files/aopa/home/training-and-safety/nall-report/20162017accidentscorecard.pdf.

AOPA Air Safety Institute (2018b). 27th Joseph T. Nall Report—General Aviation Accidents in 2015. Retrieved from https://www.aopa.org/-/media/files/aopa/home/training-and-safety/nall-report/27thnallreport2018.pdf.

Hunter, D. R. (2002). Risk perception and risk tolerance in aircraft pilots. Retrieved from https://ntlrepository.blob.core.windows.net/lib/19000/19800/19856/PB2003100818.pdf.

O'Hare, D. (1990). Pilot's perception of risk and hazard in General Aviation. *Aviation Space and Environmental Medicine, 61(7),* 599-603.

Rantz, W. G., Dickinson, A. M., Sinclair, G. A., & Van Houten, R. (2009). The effect of feedback on the accuracy of checklist completion during instrument flight training. *Journal of Applied Behavior Analysis, 42,* 497-509.

Rao, A. H., & Marais, K. (2016). Comparing hazardous states and trigger events in fatal and non-fatal helicopter accidents. *16th AIAA Aviation Technology, Integration, and Operations Conference*. Washington, DC: AIAA AVIATION.