

11-14-2014

Reliability of Eyewitness Reports to a Major Aviation Accident

Dave English

Arizona State University, dave.english@asu.edu

Michael Kuzel

Arizona State University, MICHAEL.KUZEL@asu.edu

Follow this and additional works at: <https://commons.erau.edu/ijaaa>

 Part of the [Cognition and Perception Commons](#), [Criminology and Criminal Justice Commons](#), [Forensic Science and Technology Commons](#), [Industrial and Organizational Psychology Commons](#), and the [Management and Operations Commons](#)

Scholarly Commons Citation

English, D., & Kuzel, M. (2014). Reliability of Eyewitness Reports to a Major Aviation Accident. *International Journal of Aviation, Aeronautics, and Aerospace*, 1(4). <https://doi.org/10.15394/ijaaa.2014.1040>

This Article is brought to you for free and open access by the Journals at Scholarly Commons. It has been accepted for inclusion in International Journal of Aviation, Aeronautics, and Aerospace by an authorized administrator of Scholarly Commons. For more information, please contact commons@erau.edu, wolfe309@erau.edu.

Dr. Percy Walker, director of Britain's Ministry of Aviation accident inspector branch in the early 1960's, was said by *The Sunday Times* to have researched more crashes than anybody else in the world (Air Correspondent, 1962). In the same article he was quoted as saying that eyewitnesses to aviation accidents are "almost always wrong" (p. 8). Contemporary accident investigation textbooks employ more measured language (Strauch, 2002; Wood & Sweginnis, 2006) but they do note that inconsistencies are often found among eyewitness accounts. In the 50 years since Dr. Walker's statement, research into eyewitness testimony has advanced considerably; however there remains a paucity of published empirical studies regarding the validity and reliability of aviation crash witness statements.

We have long known eyewitness testimony to be less than completely reliable (Loftus 1996; Toglia, Reed, Ross & Lindsay, 2006). Over a century ago Münsterberg (1908) gave many examples, including the time a revolver was fired during a lecture. The dramatic scene was all play-acting, part of a controlled experiment. Similar experimental techniques are still used in eyewitness research, but have limitations (Memon, Mastroberardino & Fraser, 2008). Field and archival techniques further expand our understanding (Wells, Memon & Penrod, 2006), and have included studies of witness records of large traumatic events such as the *Titanic* sinking (Riniolo, Loledin, Drakulic & Payne, 2003), the Oklahoma City Federal Building bombers (Memon & Wright, 1999) and the 9/11 attacks (Altmann, 2003). But despite over 100 years of research, and more than 2,000 papers published on eyewitness identification in the past 30 years, studies on aircraft accident eyewitness reports remain scarce. Dodge (1983) found variability in witness accounts of a major aviation crash, but the sample size ($n = 20$) was relatively small and all witnesses were on-board survivors of the crash they were describing.

There are many variables known to influence eyewitness accuracy (Brewer & Wells, 2011). Memory of an event can be changed by exposure to misinformation about the event (Zaragoza & Lane, 1994), by reactivating the memory (St. Jacques & Schacter, 2013), by trauma and perceived culpability (Foster & Naylor, 1999), by social conformity (Edelson, Sharot, Dolan & Dudai, 2011), and by talking about the event (Wright, Memon, Skagerberg & Gabbert, 2009). However people evaluating the statements are often not aware of these issues. A review of eyewitness reliability in motor vehicle crashes (Robins, 2009) notes that juries show a "marked preference" for eyewitness evidence over what should be the more compelling physical evidence. A survey of potential jurors in the District of Columbia (Schmechel, O'Toole, Easterly & Loftus, 2006) finds they "misunderstand how memory generally works and how particular factors ...

affect the accuracy of eyewitness testimony” (p. 194). Even informed psychologists may be misled. Chan, Thomas and Bulevich (2009) show “real-life eyewitness memory may be even more susceptible to misinformation than is currently envisioned” (p.66).

Following a summary of several aircraft accidents that seem to provide some qualitative support to claims of mistaken eyewitness accounts, we subject the extensive archival witness record of a major aviation accident to statistical analysis. The aim of the study, to demonstrate the reliability of eyewitness reports to a major aviation accident, is achieved with several statistical analyses converging towards a conclusion of unreliability.

Qualitative Examples of Aviation Accident Eyewitness Validity and Reliability

There are many aircraft accident reports where eyewitnesses are generally in agreement with each other and the final accident probable cause; for example the takeoff of a DC-9 seen by 100 external observers where “none of the witnesses described smoke or flames coming from any part of the airplane other than the right engine” (National Transportation Safety Board, 1987). Sometimes a single eyewitness can supply otherwise ephemeral evidence, as for example the farmhand who reported that something fell off an accident aircraft: “Whizzed past me [h]ead it did, and when I dug it out of ground a large chunk of ice it were” (Brown, 1962, p. 38). The probable cause of the crash was determined to be inflight icing based largely on the farmhand’s account. However there are also many well-recorded cases that support the late Dr. Walker’s contention, cases where aviation accident eyewitnesses report seeing things that did not happen or substantially confuse the order of events.

At an airshow in 1952, a supersonic fighter disintegrated in the air causing the death of both crew and 29 spectators (Staff, 1952). Over 100,000 people witnessed the accident. A public appeal was put out for witness accounts and photographs to help solve the mystery, resulting in several thousand letters being collected. Rivas and Bullen (2008) found “many of the accounts are touchingly detailed and well intentioned, but the whole of the vast mail was of little use” (p. 186). The vital clue that led to determination of probable cause was supplied by a cine film. The in-flight breakup happened in less than a second, and almost all the eyewitnesses, including experienced pilots, gave grossly inaccurate accounts when compared to the film record.

Investigating the mid-air collision of a passenger DC-9 and a Marine Corps F-4 the National Transportation Safety Board (NTSB) found “witnesses in the area of the accident gave widely varying accounts” (NTSB, 1972). Five people described both aircraft on steady courses prior to the collision, but fifteen people described the fighter aircraft in a rolling or evasive maneuver prior to the collision.

Wilkinson (1977) quotes an eyewitness to a crash describing a light aeroplane just before impact as “heading right toward the ground—straight down” (p.102). However photographs of the crash site clearly showed the aeroplane plane hit flat and at a low enough angle to skid for almost 1,000 feet. Two expert eyewitnesses to a crash on takeoff of a MD-82 stated that the wing flaps were extended, but the Board determined the flaps were in fact not extended (NTSB, 1988). What was initially reported as a possible bombing of a B767 due to many eyewitness accounts of the plane first exploding in fire (Johnson, 1991; Kelly & Elliott, 1991) turned out to be caused by the uncommanded activation of an engine thrust reverser (Aircraft Accident Investigation Committee, 1993).

When a fuel tank explosion caused a B747 to descend in pieces from 13,000 feet, the fireball was seen by hundreds of people, about one-third of who reported that they observed a streak of light moving upward in the sky (NTSB, 2000). However there was no evidence that a missile struck the plane, and physical examination of the wreckage unequivocally supports the cause as a fuel tank explosion. Thirty-eight of the witnesses described a streak of light as ascending vertically. Forty-five reported that a streak moved to the east, 23 that it moved to the west, 18 that it moved to the south, and 4 that it moved to the north.

When a MD-82 crashed on takeoff initial reports included eyewitness accounts of an engine catching fire as the aeroplane heading down the runway (Goodman, Todd & Koch 2008; Naughton & Strange, 2008). However analysis showed that engine performance was normal on takeoff, and that the cause of the crash was failure to set the flaps (CIAIAC, 2008).

A Cessna 310R that crashed following a steep descent was observed by several witnesses. All described to the NTSB (2012) a nose-down, vertical descent to ground contact. However some described the engine sound as “full throttle,” “wide open,” “really loud,” and “never let up on [the] throttle” But others stated the engine was “puttering” or “quit” before the descent (para. 3). One witness believed he had seen a meteorite. A veteran aviation journalist wrote of the eyewitness accounts, “as is often the case, they disagreed” (McClellan, 2013, p. 84).

Considering these examples, it's understandable that not just journalists are unconvinced by accident eyewitness accounts. NTSB Member John Goglia declared that eyewitness reports of aeroplane crashes are often mistaken (Halbfinger & Wald, 2003), and NTSB spokesperson Ted Lopatkiewicz asserted "eyewitness testimony at a plane crash ... is unreliable" (Wald, 2002, p. 5). These professional opinions are not however shared by the general public (Simons & Chabris, 2011). Eyewitness reports are often given prominence in press reports and used in legal actions. Philosophers have long grappled with the dual ideas of human conviction and disagreement, with Renouvier declaring in 1859 that "properly speaking, there is no certainty; there are only people who are certain" (Schulz, 2010, p. 159). Decades of experience have led aviation crash investigators to discount eyewitness accounts, and instead rely on flight recorders, radar recordings and physical evidence to determine cause. There are, however, no systematic published studies of aviation crash eyewitness accounts to validate investigators beliefs. This is due in part to the problems of creating a controlled experiment. It is much easier for a psychology lab to stage a fistfight or a robbery, than it is to stage a burning airliner coming out of the sky.

Description of AA587 Crash

On the morning of November 12th, 2001, an American Airlines wide-body A300 jet climbed into clear sunny skies over New York City on what was planned to be a routine flight. A few minutes later flight 587 violently pitched down and crashed into the middle of a residential neighborhood (Kleinfeld, 2001). All 260 people on board died, along with an additional five people on the ground. Coming two months after the 9/11 attacks, there were initially fears that this may have been another terrorist attack. A large NTSB and Federal Bureau of Investigation (FBI) effort was mounted to determine what happened, an investigation that included interviewing hundreds of eyewitnesses.

The NTSB determined that the probable cause of flight 587's accident was co-pilot rudder inputs that resulted in the vertical stabilizer breaking off from the fuselage and falling into Jamaica bay (NTSB, 2004). The loss of a primary aerodynamic flight control surface and substantial change in center of mass caused the jet to pitch down and dive towards the ground. Recovery was impossible. Fifteen seconds after rudder separation the plane crashed into the quiet Belle Harbor suburb of Queens, New York, a three-kilometer (1.8 mile) by one kilometer (0.6 mile) (at its widest) neighborhood located on a barrier peninsula between Jamaica Bay and the Atlantic Ocean (figure 1). The majority of the aircraft made landfall close to the geographic center of the community, destroying three homes and damaging six others in a post-impact fire. Both

engines separated from the wings before impact, landing about 200 metres from the main wreckage. The accident was not survivable for the airplane occupants.



Figure 1. Photograph of Belle Harbor. Taken from an altitude of 2000 feet over the Atlantic Ocean, looking northwest with Jamaica Bay and New York City in the background; it was into the center of this neighborhood that American Airlines 587 crashed. (Photo by author D. English.)

The weather that morning was clear and brisk. Immediately before the accident JFK airport reported the wind to be from 310° at 11 knots, visibility unrestricted, a few clouds at 4,300 feet with a temperature of 6° C. The sun was positioned 22.5° above the horizon, bearing 142° true (NTSB, 2004). *The Times* called it a brilliant, blue sky (Kleinfield 2001). Belle Harbor residents are under a busy flight path for JFK departures, so they are used to seeing airliners pass overhead. Flight 587 was determined by the NTSB to have been viewed by a total of 354 witnesses that provided sufficient detail to document. More pertinently, the majority of witnesses were concentrated in Belle Harbor. It is extremely rare to have such a numerous, compact group unexpectedly observe a nearby low-altitude airliner crash in good weather from all angles. The fact that these eyewitnesses must have had an emotional reaction to the disaster, and no doubt later watched TV news reports, and/or read newspaper accounts of the accident, and discussed their observations with others, before making their statements might seem to diminish their use in eyewitness research. It is known that human memory is strikingly susceptible to social (Wright, Memon, Skagerberg & Gabbert, 2009) and other sullyng influences (Foster & Naylor 1999; Robins, 2009). But since we are concerned with real world accident investigations this seeming contamination is the more expected condition, the reality of actual witness statements, and maybe gives archival methodology more validity than experimental techniques (Christianson, 1992).

Materials and Methods

Recording witness statements was earnest work due to the number of fatalities and the immediate lack of a clear cause. Interviews were undertaken by a large number of NTSB, local police and FBI agents. Using the NTSB Witness Group Factual Report (Schiada, 2002) as a model, we created an Excel spreadsheet containing all 354 witness summaries. While there are considerable original records of the interviews, it was not recorded how the questions were posed, or what other reports (media/social) may have influenced the witness. Our dataset used only the original consensus NTSB Witness Group interpretation of each witness's testimony.

Our goal in constructing a dataset to analyze was to have the highest likelihood of finding reliability. By removing obvious outliers and controlling for geographic position we stood the best chance of discovering patterns. And if no patterns are found in this clean dataset, we strengthen the hypothesis that data from eyewitness reports are unreliable. We removed witnesses that were moving on boats, trains, planes or whose position was otherwise uncertain. We also removed witnesses who were many kilometers from the crash site, resulting in a dataset containing 239 Belle Harbor witnesses (defined as all ground witness between 149th Street and 108th Street). Coding the witness location into the dataset would allow us to control for position relative to the crash. We determined the geographic coordinates of the Belle Harbor witnesses using the information reported in the NTSB Appendixes A and B (street address or textual description of location) in combination with an examination of Google Earth and Google Street View. The resulting Keyhole Markup Language (KML) file is rendered as figure 2 using the USGS National Map Viewer. Location was initially coded into the spreadsheet using a city-block grid centered on the crash site at 131st Street and Newport Avenue. We then performed a matrix rotation, giving us witness location using Cartesian coordinates with origin at main crash site, abscissa as position left or right of aircraft track, and ordinate as position along aircraft track. We also converted the Cartesian coordinates into polar coordinates for analysis based on distance from the crash site. The data were now in a useful format to allow us to perform a series of statistical measures on the witness statements.



Figure 2. Belle Harbor Eyewitness Locations. The gray line ending in a red marker shows the path of the main body of the aircraft. Each green marker is an eyewitness, the other red markers show final location of the engines and left wing tip.

Analysis 1: Fire

Of the 239 Belle Harbor witnesses 183 expressed an opinion about the airliner being on fire or not before it impacted the ground. The NTSB stated that the no-fire code was only “utilized in situations when a witness specifically indicated he or she did not see any fire” (Schiada, 2002, p. 4). This leaves 56 (23%) that expressed no opinion on this question. Our analysis used only those witnesses that expressed a definite opinion, in a continuing attempt to collect the cleanest, most reliable, witness pool.

Belle Harbor witnesses were not in agreement about the airliner being on fire or not before it impacted the ground. The distribution of the 183 witnesses that reported either observing no fire (65) or some fire (118) was investigated using RStudio 0.97.311 (R Core Team, 2012) and MATLAB 2011a (MATLAB,

2011) software with all tests of significance at a $p = .05$ criterion. The majority opinion of all eyewitnesses expressing an opinion on pre-impact fire is statistically significant (using chi-square test with an equally divided fire/no-fire null hypothesis ($\chi^2 [1, N = 183] = 15.35, p < .001$). Graphical mapping shows little discernable geographic pattern (figure 3). Separate logistic regression analyses with fire/no-fire as the binary categorical dependent variable and a geographic dimension (in km) as the independent (predictor) variable reveals no significant relationship along the abscissa ($b = 0.39, p = .109$), the ordinate ($b = 0.14, p = .788$), or distance from the crash site ($b = -0.22, p = .567$). This confirms the intuitive conclusion from the graphical mapping that there is no preferred orientation for these eyewitness reports. Indeed, the overall distribution of the fire witnesses compared to no-fire witnesses appears by eye to be about the same. This more general assertion is harder to test, as the witness pool was drawn from a decidedly non-uniformly distributed population (on a peninsula, tending to be on busy streets or popular locations).

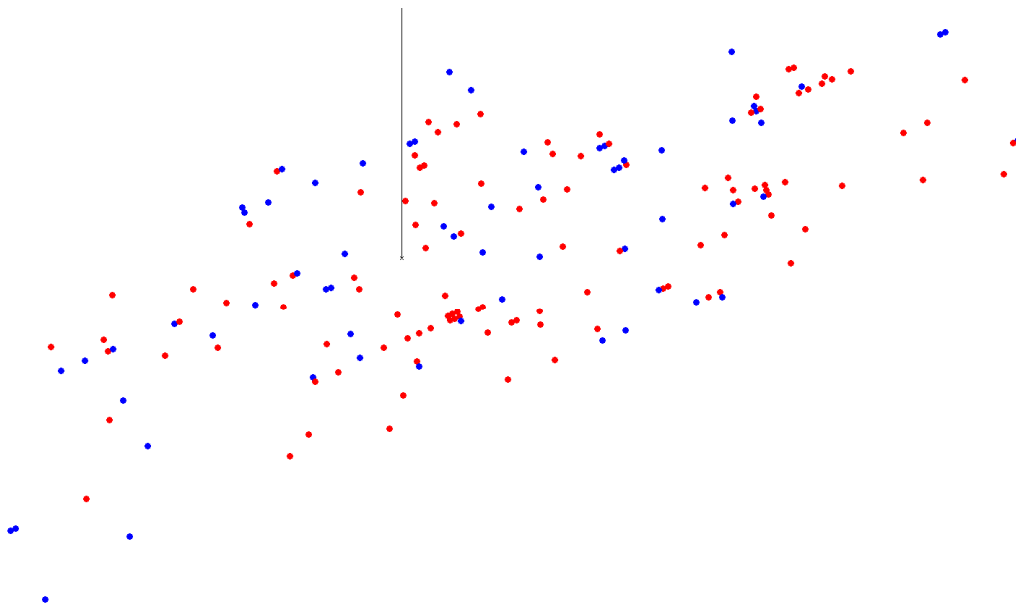


Figure 3. Was there pre-impact fire? Witnesses that reported seeing pre-impact fire are in red, those that saw no fire in blue. The gray line is the flight path. Some locations jittered to prevent overplotting.

We investigated independence with a two-tailed two-dimensional paired Kolmogorov-Smirnov test. Peacock's algorithm (Peacock, 1983) determines whether two sets of data arise from the same or different distributions without making any assumptions of the distribution. While the Kolmogorov-Smirnov test has sometimes been considered a weak form of analysis, it has the required advantage here that it is non-parametric and distribution free. Use of Peacock's test for multi-dimensional data sets is well established (Lopes, Reid & Hobson, 2007). The null hypothesis is that both data sets are drawn from the same continuous distribution. For the fire/no fire distributions, the null hypothesis is not rejected ($D = 0.167$, $p = 0.200$), suggesting that indeed the fire and no-fire distributions are the same.

Assuming the distribution of fire/no-fire witnesses to be uniform, we can calculate the results of polling smaller samples of witnesses from the group without regard to location. While the majority opinion is statistically significant, the variance is considerable. This becomes increasingly important when considering the small number of witnesses that view most aviation accidents. Applying hypergeometric distribution analysis to the dataset values, a poll of 5 witnesses reveals a 24% probability that a majority of such a sample would have reported seeing no pre-impact fire (the opposite of the majority opinion in the population). With 10 witnesses there is still a 9% probability that a majority would report no fire, and a 92% probability that 2 of the 10 would report against the population majority opinion. Clearly the variance found in this compact eyewitness population means that eyewitness consensus in commonly sized witness samples is unlikely and that a majority of such a sample may disagree with the population majority opinion.

As to location of the fire, Belle Harbor observers that saw fire offered varying details: 7% said there was a fire in the right engine, 7% in the right wing, 6% in the tail, 41% in the fuselage, 9% in the left engine, 17% in the left wing, 14% in a miscellaneous area, 4% in an undefined wing, and 4% in an undefined engine. The NTSB determined "witnesses who reported observing the airplane on fire were most likely observing a fire from the initial release of fuel or the effects of engine compressor surges" (NTSB, 2004). The engines themselves suffered no in-flight fire, and there was probably no other pre-impact fire.

Analysis 2: Smoke

Belle Harbor witnesses were not in agreement about the airliner leaving a smoke trail or not before it impacted the ground. The distribution of the 105 witnesses that reported either no smoke (60) or some smoke (45) was investigated

using the same tools and methods. Again, graphical mapping shows no discernable geographic pattern (figure 4). The majority opinion of no smoke is not statistically significant using a chi-square test with a no preference null hypothesis ($\chi^2 [1, N = 105] = 2.14, p = .143$). Logistic regression analyses show no significant relationship to the smoke/no smoke reports along the abscissa ($b = 0.30, p = .296$) or the ordinate ($b = -0.14, p = .818$). There is an observed trend for an increased probability to report smoke with increasing distance from the crash site, but this was not statistically significant ($b = 0.93, p = .066$). A two-tailed two-dimensional paired Kolmogorov-Smirnov test does not reject the null hypothesis that the two data sets are drawn from the same continuous distribution ($D = 0.239, p = 0.200$).

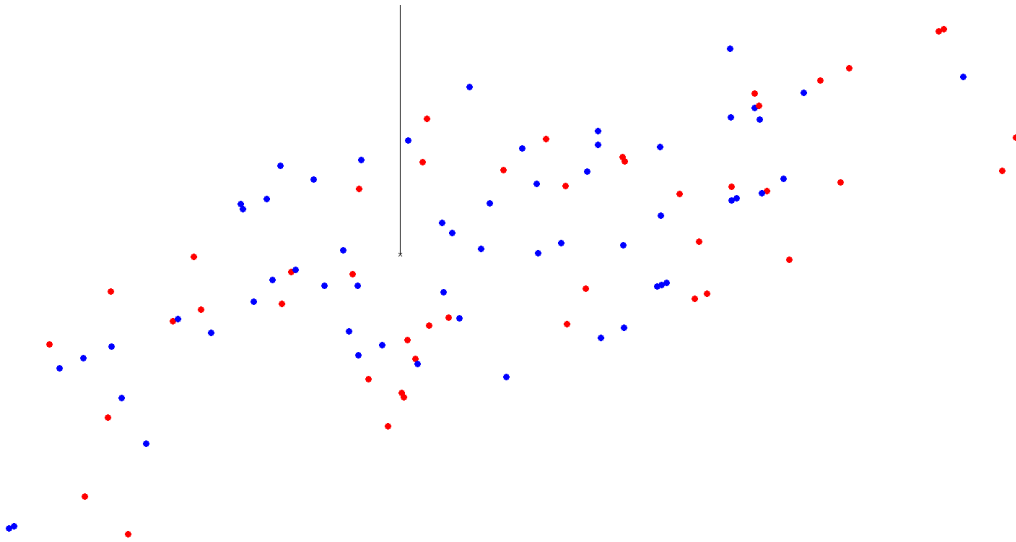


Figure 4. Was there pre-impact smoke? Witnesses that reported seeing pre-impact smoke in red, those that reported no smoke in blue. The gray line is the flight path.

Hypergeometric distribution analysis reveals that if five random witnesses were polled from the group, there is a 36% probability that a majority of them would have reported pre-impact smoke, the opposite of the no smoke majority with this population. With 10 witnesses there is still a 21% probability that a majority would report smoke, and a 97% probability that 2 of the 10 would report against the population majority opinion.

As to location of the pre-impact smoke, Belle Harbor observers that saw smoke offered varying details: 2% reported the right engine, 13% the right wing, 7% the tail, 29% the fuselage, 9% the left engine, 7% the left wing, 42% from a miscellaneous area, 4% an undefined wing, and 2% an undefined engine. The NTSB report reached no conclusion about how much or from where the jetliner was emitting smoke prior to impact.

Analysis 3: Noise

The distribution of the 239 witnesses that reported either in-flight noise (156) or no noise (83) was investigated using the same tools and methods. Graphical mapping shows no discernable geographic pattern (figure 5). Logistic regression analyses show no significant relationship to the noise/no noise reports along the abscissa ($b = -0.02$ $p = .287$) or the ordinate ($b = 0.01$ $p = .831$). There is an observed trend for a decreased probability to report noise with increasing distance from the crash site; with about 70% of witnesses immediately adjacent to the crash site reporting hearing noise, decreasing to about 50% reporting noise when two kilometers (1.2 miles) away. However this was not statistically significant ($b = -0.04$ $p = .105$). A two-tailed two-dimensional paired Kolmogorov-Smirnov test does not reject the null hypothesis that the two data sets are drawn from the same continuous distribution ($D = 0.210$ $p = 0.200$).

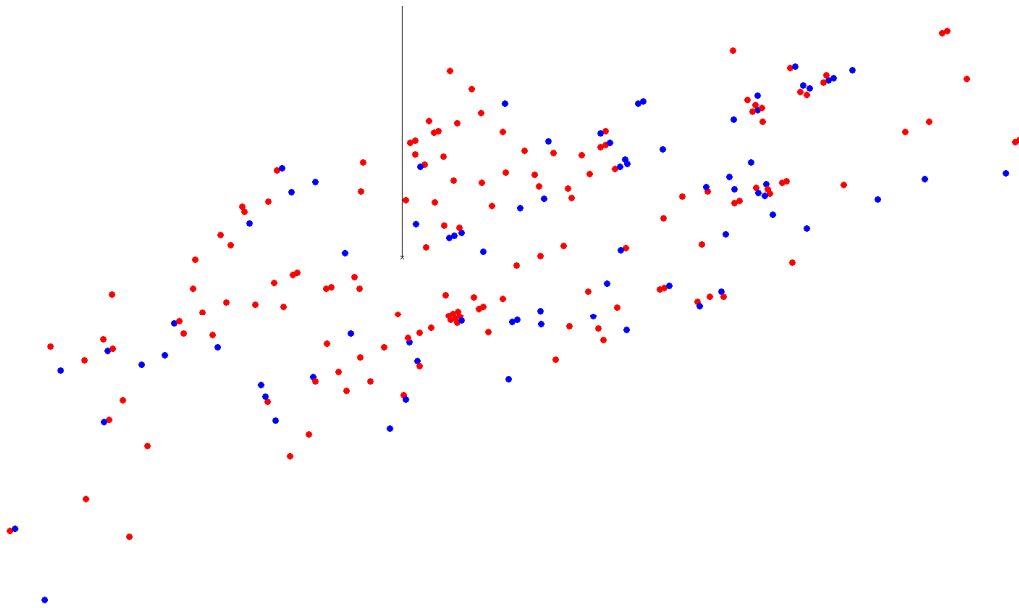


Figure 5. Was there pre-impact noise? Witnesses that reported hearing pre-impact noise in red, those that reported no noise in blue. The gray line is the flight path.

The majority opinion of in-flight noise is statistically significant using a chi-square test with a no preference null hypothesis, ($\chi^2 (1, N = 239) = 22.30, p < .001$). Hypergeometric distribution analysis reveals that if five witnesses were randomly polled, there is a 23% probability that a majority of them would have reported no in-flight noise, the opposite of the in-flight noise majority with this population. With 10 witnesses there is still a 9% probability that a majority would report no noise, and a 92% probability that 2 of the 10 would report against the population majority opinion.

Analysis 4: Explosion

Forty-one (17%) of the Belle Harbor witness pool reported that in addition to noise, they heard a pre-impact explosion (or boom or loud pop). Geographical mapping shows no discernible pattern (figure 6), confirmed by logistic regression analyses that show no statistically significant relationship to the explosion/no-noise reports along the abscissa ($b = -.03, p = .282$), the ordinate ($b = -.68, p = .228$) or distance from the crash site ($b = 0.30, p = .486$). A two-tailed two-dimensional paired Kolmogorov-Smirnov test does not reject the null hypothesis that the two data sets are drawn from the same continuous distribution ($D = 0.182, p = 0.200$).

The majority opinion of no explosion is statistically significant using a chi-square test with a no preference null hypothesis ($\chi^2 [1, N = 239] = 175.84, p < .001$). Hypergeometric distribution analysis reveals that if five witnesses were polled from the group, there is a 4% probability that a majority of them would have reported a pre-impact explosion, the opposite of majority opinion with this population. This is a much lower probability than the other categories studied, but even in this case the probability of all five witnesses agreeing there was no explosion is only 61%. With 10 witnesses there is a 1.5% probability that a majority would report an explosion, and a 54% probability that 2 of the 10 would report against the population majority opinion.

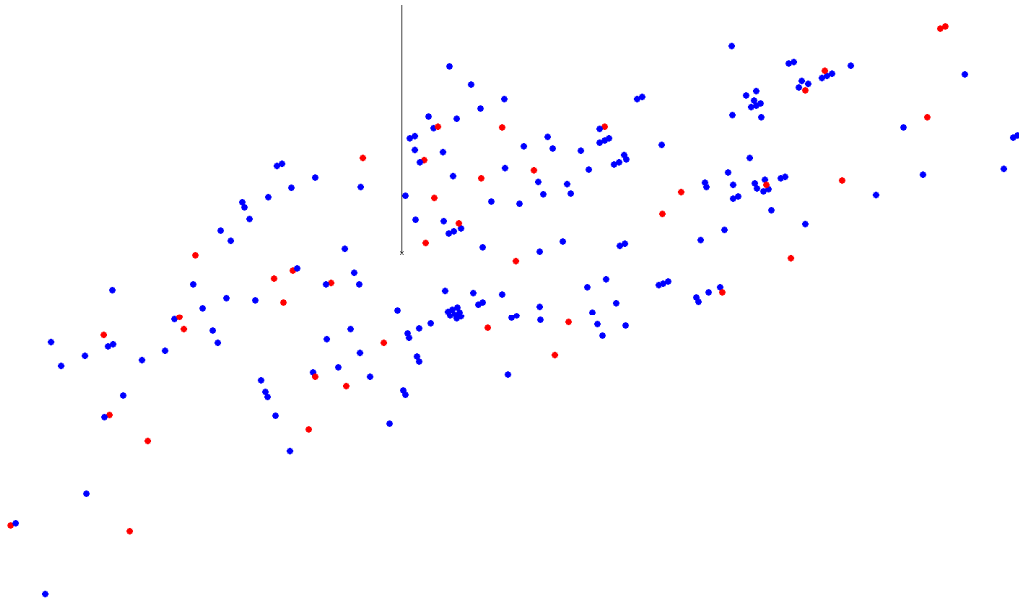


Figure 6. Was there a pre-impact explosion? Witnesses that reported a pre-impact explosion in red, those that reported no explosion in blue. The gray line is the flight path.

Limitations

This study is subject to all the limitations inherent in archival eyewitness research. This includes, but is not limited to, a limited sample size, a nonrandom sample, participant self-selection, no control for retention interval, eyewitnesses with varied vantage points, conflation of geographic position with other factors, multicollinearity, lack of control of eyewitness interviews, no control for emotional state, no control for post-observation social or media influence, possible limited generalizability of one specific type of accident to other aviation accidents, and an inability to manipulate variables. Interviews were conducted by three different organizations (local police/FBI/NTSB) at varying times using varying formats (written submissions/telephone/face-to-face) with no records of witness media or social exposure. The NTSB found disagreements between statements given to different investigators at different times (Schiada, 2002). And they noted that, “disagreement and direct conflicts also existed between statements from the same source (e.g. two or more police statements pertaining to the same witness)” (p.7).

Despite the methodological limitations, many authors have supported archival research as a naturalistic contribution to the use of multiple investigational techniques to uncover reality (Christianson, 1992; Christianson, Goodman & Loftus, 1992).

Discussion

Eyewitness reports to this airplane crash show considerable disparities. Even with over two hundred witnesses in a compact geographic area stating an observation, the variance was large enough to preclude forming a statistically significant conclusion about a basic large-scale event (was the aircraft trailing smoke). And while in three other cases a statistically significant conclusion could be reached from the large 'clean' dataset, the observed variances were still considerable. These quantitative results are in agreement with the apparent variability of eyewitness statements qualitatively reported for other widely observed aviation accidents.

Furthermore, analysis of the geographic distribution of the witness group observations shows no significant pattern. Logistic regression showed no statistically significant structure for four types of observation along three possible dimensions of regularity. Two-tailed two-dimensional paired Kolmogorov-Smirnov tests showed that eyewitnesses reporting opposite observations were drawn from the same continuous distribution. We might expect some directions relative to the flight path of the aircraft to give a clearer view of an airborne event, or people closer to an event to be better (or worse) witnesses, but our findings show that in this case location of a witness does not appear to significantly affect the validity of major observations reported. The variance was evenly spread by geographic location within the witness pool.

If the accident were to have been observed by only a small group of witnesses (as is common), there is considerable probability that the witnesses would not agree; and there is the possibility that a majority of the witnesses would report against the population majority. That small groups have a remarkably high chance of not following the population characteristics has been proven since Pascal's Triangle in the 1600's (Mlodinow, 2008); and the mistaken intuition that a small sample will accurately reflect underlying probabilities was called the law of small numbers by Kahneman and Tversky (1971). Thus, the findings for commonly sized small samples are not surprising. However, the inability to make a statistically significant conclusion about a basic large-scale event from the large ($N > 200$) sample is unexpected.

Combining the high variances, the lack of any preferred observational perspective and the law of small numbers, a conclusion of eyewitness unreliability for aviation crashes is unavoidable. Although they sound compelling — “*I saw the crash with my own eyes*” — a small group of witnesses to an aviation accident giving reports several days after the event may well not produce reliable information. This is demonstrated to be true for a crash as seemingly perceptible as a wide-body transport jet at low altitude in clear daylight conditions. Dr. Percy Walker’s claim that eyewitnesses to aviation accidents are almost always wrong is certainly not proven. But the current reported practice by accident investigators of placing low value to eyewitness accounts of aircraft crashes is supported by the empirical evidence.

References

- Air Correspondent. (1962, March 11). 30 men who wait for an air crash. *The Sunday Times*, London, UK, p. 8.
- Aircraft Accident Investigation Committee (1993). *Lauda Air Luftfahrt Aktiengesellschaft, Boeing 767-300ER, registration QE-LAV, Dan Chang district, Suphan Buri province, Thailand, 26 May B.E. 2534 (A.D. 1991)*. Thailand: Ministry of Transport and Communications.
- Altmann, E. M. (2003). Reconstructing the serial order of events: a case study of September 11, 2001, *Applied Cognitive Psychology*, 17, 1067-1080. <http://dx.doi.org/10.1002/acp.986>
- Brewer, N. & Wells, G. L. (2011). Eyewitness identification. *Current Directions in Psychological Science*, 20, 24-27. <http://dx.doi.org/10.1177/0963721410389169>
- Brown, V. (1962, September). Accident aftermath. *Aviation and Space Magazine*, 38-39.
- Chan, J. C. K., Thomas, A. K., & Bulevich, J. B. (2009). Recalling a witnessed event increases eyewitness suggestibility. *Psychological Science*, 20(1), 66-73. <http://dx.doi.org/10.1111/j.1467-9280.2008.02245.x>
- Christianson, S. (1992). Emotional memories in laboratory studies versus real-life studies: Do they compare? In M. A. Conway, D. C. Rubin, H. Spinnler, & W. A. Wagenaar (Eds.), *Theoretical perspectives on autobiographical memory* (pp. 339-353). Dordrecht, The Netherlands: Kluwer. http://dx.doi.org/10.1007/978-94-015-7967-4_20
- Christianson, S., Goodman, J., & Loftus, E. F. (1992). Eyewitness memory for stressful events: Methodological quandaries and ethical dilemmas. In S. Christianson (Ed.), *The handbook of emotion and memory: Research and theory* (pp. 217-241). Hillsdale, NJ: Erlbaum.
- CIAIAC. (2008). *Interim report A-032/2008. Accident involving aircraft McDonnell Douglas DC-9-82 (MD-82), registration EC-HFP, operated*

by Spanair, at Madrid-Barajas airport on 20 August 2008. Madrid, Spain: Ministerio de Fomento.

Dodge, R. E. (1983). Aircraft accident survivors as witnesses. *Aviation, Space, and Environmental Medicine*, 54(2), 165-167.

Edelson, M., Sharot, T., Dolan, R. J., & Dudai, Y. (2011). Following the crowd: Brain substrates of long-term memory conformity. *Science*, 333, 108-111. <http://dx.doi.org/10.1126/science.1203557>

Foster, A. R., & Naylor, K. (1999). Trauma and perceived culpability as factors in eye-witness accounts. *Proceedings of the 4th International Conference of The Institute of Traffic Accident Investigators*. Telford, UK. 73-79.

Goodman, A., Todd, B., & Koch, K. (2008, August 20). *Passengers reportedly saw flames before jet crashed, killing 153*. Retrieved from <http://edition.cnn.com/2008/WORLD/europe/08/20/plane.crash.madrid/index.html>

Halbfinger, D. M., & Wald, M. L. (2003, January 9). 21 die in commuter plane crash in North Carolina. *The New York Times*.

Johnson, S. S. (1991, May 29). Scavengers complicate crash probe. *USA Today*, p. 4A.

Kelly, N., & Elliott, H. (1991, May 28). Baggage bomb is blamed for Thai air crash. *The Times*.

Kleinfield, N. R. (2001, November 13). The crash of flight 587: The overview. *The New York Times*.

Loftus, E. F. (1996). *Eyewitness testimony: With a new preface by the author*. Cambridge, MA: Harvard University Press.

Lopes, R. H. C., Reid, I., & Hobson, P. R. (2007). The two-dimensional Kolmogorov-Smirnov test. *XI International Workshop on Advanced Computing and Analysis Techniques in Physics Research*, Amsterdam, Netherlands.

MATLAB. (2011). *MATLAB and Statistics Toolbox Release R2011a*. Natick, MA: The Mathworks Inc.

- McClellan, J. M. (2013). What you don't know can . . . *Sport Aviation*, 62(1), 84-86.
- Memon, A., Mastroberardino, S., & Fraser, J. (2008). Münsterberg's legacy: What does eyewitness research tell us about the reliability of eyewitness testimony? *Applied Cognitive Psychology*, 22, 841-851.
- Memon, A., & Wright, D. B. (1999). Eyewitness testimony and the Oklahoma bombing. *Psychologist*, 12(6), 292-295.
- Mlodinow, L. (2008). *The drunkard's walk: How randomness rules our lives*. New York, NY: Pantheon.
- Münsterberg, H. (1908). *On the witness stand: Essays on psychology and crime*. Garden City, NY: Doubleday, Page & Company.
- National Transportation Safety Board. (1972). *Aircraft accident report, Hughes Air West DC-9, N9345, and U.S. Marine Corps F-4B, 151458, near Duarte, California, June 6, 1971*. NTSB-AAR-72-26. Washington DC: National Transportation Safety Board.
- National Transportation Safety Board. (1987). *Aircraft accident report, Midwest express airlines, inc., DC-9-14, N100ME, general Billy Mitchell field, Milwaukee, Wisconsin, September 6, 1985*. NTSB/AAR-87/01. Washington, DC: National Transportation Safety Board.
- National Transportation Safety Board. (1988). *Aircraft accident report. Northwest airlines, inc., McDonnell Douglas DC-9-82, N312RC, Detroit metropolitan wayne county airport, Romulus, Michigan, August 16, 1987*. NTSB/AAR-99/05. Washington, DC: National Transportation Safety Board.
- National Transportation Safety Board. (2000). *Aircraft accident report. In-flight breakup over the Atlantic ocean, Trans World Airlines flight 800, Boeing 747-131, N93119, near east moriches, New York, July 17, 1996*. NTSB/AAR-00/03. Washington, DC: National Transportation Safety Board.
- National Transportation Safety Board. (2004). *Aircraft accident report, in-flight separation of vertical stabilizer, American Airlines flight 587, Airbus Industrie A300-605R, N14053, Belle Harbor, New York, November 12,*

2001. NTSB/AAR-04/04. Washington, DC: National Transportation Safety Board.
- National Transportation Safety Board. (2012). Probable cause report ERA11FA185. Retrieved from http://www.nts.gov/aviationquery/brief2.aspx?ev_id=20110311X65739&ntsbno=ERA11FA185&akey=1
- Naughton, P., & Strange, H. (2008, August 20). 152 burnt to death in Madrid plane disaster. *The Times*.
- Peacock, J.A. (1983). Two-dimensional goodness-of-fit testing in astronomy. *Monthly Notices Royal Astronomy Society*, 202, 615-627. <http://dx.doi.org/10.1093/mnras/202.3.615>
- R Core Team. (2012). R: A language and environment for statistical computing. R Foundation for statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org/>.
- Robins, P. J. (2009). *Eyewitness reliability on motor vehicle crashes: A primer for practitioners*. (2nd ed). Tucson, AZ: Lawyers & Judges Publishing Company.
- Riniolo, T. C., Koledin, M., Drakulic, G. M., & Payne, R. A. (2003). An archival study of eyewitness memory of the Titanic's final plunge. *Journal of General Psychology*, 130(1), 89-95. <http://dx.doi.org/10.1080/00221300309601277>
- Rivas, B., & Bullen, A. (2008). *John Derry: The story of Britian's first supersonic pilot* (New Edition ed.). Yeovil, UK: Haynes Publishing.
- St. Jacques, P. L., & Schacter D. L. (2013). Modifying memory: Selectively enhancing and updating personal memories for a museum tour by reactivating them. *Psychological Science*, 24(4), 537-542. <http://dx.doi.org/10.1177/0956797612457377>
- Schiada, L. (2002). *Witness group factual report: DCA02MA001*. Washington, DC: National Transportation Safety Board.
- Schmechel, R. S., O'Toole, T. P., Easterly, C. & Loftus, E. F. (2006). Beyond the ken? Testing jurors' understanding of eyewitness reliability evidence.

- Jurimetrics, The Journal of Law, Science, and Technology*, 46(2), 177-214.
- Schulz, K. (2010). *Being wrong: Adventures in the margin of error*. New York, NY: Ecco Press.
- Simons, D. J., & Chabris, C. F. (2011). What people believe about how memory works: A representative survey of the U.S. population. *PLoS ONE*, 6(8), e22757. <http://dx.doi.org/10.1371/journal.pone.0022757>
- Staff. (1952, September 12). The Farnborough tragedy. *Flight and Aircraft Engineer*, LXII(2277), p. 334.
- Strauch, B. (2002). *Investigating human error: Incidents, accidents, and complex systems*. Aldershot, UK: Ashgate.
- Toglia, M P., Reed, J. D., Ross, D. F., & Lindsay, R. C. (2006). *The handbook of eyewitness psychology: Volume I: Memory for events*. London, UK: Psychology Press.
- Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76(2), 105-110.
- Wald, M. L. (2002, June 23). For air crash detectives, seeing isn't believing. *The New York Times*, Section 4, p. 5.
- Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest*, 7(2), 45-75.
- Wilkinson, S. (1977, July). Charade. *Flying*, 101(1), 120.
- Wood, R. H., & Sweginnis, R W. (2006). *Aircraft Accident Investigation* (2nd ed.). Casper, WY: Endeavor.
- Wright D. B., Memon A., Skagerberg E. M., & Gabbert F. (2009). When eyewitnesses talk. *Current Directions in Psychological Science*, 18, 174-178. <http://dx.doi.org/10.1111/j.1467-8721.2009.01631.x>
- Zaragoza, M. S. & Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology*

Learning Memory and Cognition, 20, 934-945.
<http://dx.doi.org/10.1037/0278-7393.20.4.934>