Safety voice and safety listening during aviation accidents: Cockpit voice recordings reveal that speaking-up to power is not enough

Mark C Noort[ab*]

Tom W Reader[a]

Alex Gillespie[ac]

[a] London School of Economics and Political Science, The United Kingdom

[b] Leiden University, The Netherlands

[c] Bjørknes University College, Norway

[*] Corresponding author. E-mail address: m.c.noort@law.leidenuniv.nl

ABSTRACT

Safety voice is theorised as an important factor for mitigating accidents, but behavioural research during actual hazards has been scant. Research indicates power distance and poor listening to safety concerns (safety listening) suppresses safety voice. Yet, despite fruitful hypotheses and training programs, data is based on imagined and simulated scenarios and it remains unclear to what extent speaking-up poses a genuine problem for safety management, how negative responses shape the behaviour, or how this can be explained by power distance. Moreover, this means it remains unclear how the concept of safety voice is relevant for understanding accidents. To address this, 172 Cockpit Voice Recorder transcripts of historic aviation accidents were identified, integrated into a novel dataset (n = 14,128 conversational turns), coded in terms of safety voice and safety listening and triangulated with Hofstede's power distance. Results revealed that flight crew spoke-up in all but two accidents, provided the first direct evidence that power distance and safety listening explain variation in safety voice during accidents, and indicated partial effectiveness of CRM training programs because safety voice and safety listening changed over the course of history, but only for low power distance environments. Thus, findings imply that accidents cannot be assumed to emerge from a lack of safety voice, or that the behaviour is sufficient for avoiding harm, and indicate a need for improving interventions across environments. Findings underscore that the literature should be grounded in real accidents and make safety voice more effective through improving 'safety listening'.

Keywords: safety voice; safety listening; accidents; power distance; CRM.

HIGHLIGHTS

- Safety voice – the act of speaking-up about safety – is assumed to prevent harm.
- Yet, evidence from real accidents remains scant, limiting intervention design.
- In contrast to prevailing thought, flight crew spoke-up across real aviation accidents.
- This was explained by poor safety listening and high power distance.
- CRM training has improved safety voice, but only for lower power distance countries.

1. INTRODUCTION

Safety voice is the act of speaking-up about perceived hazards (1,2). For high reliability industries such as aviation, safety voice is assumed to be central to maintaining safe operations (3) and where team members withhold safety concerns ('safety silence'), or fail to engage and dismiss them (i.e., poor 'safety listening'), this has contributed to tragic accidents due to information about risk not being shared or used (4–6). Explanations for the absence of safety voice and poor safety listening during safety critical scenarios often focus on cultural norms and asymmetric leader-follower relationships (i.e., power distance (7)). Specifically, accidents are assumed to emerge from people not speaking-up due to fears for the social consequences of incorrectly raising concerns or undermining leaders (2,8–10), and poor safety listening to voice is understood to arise from norms for communication (7,11) and expected asymmetries on expertise for managing safety (12). Studies utilising vignette (13), laboratory (14), high-fidelity simulator scenarios (15) and case studies (16) have explored safety voice and safety listening extensively, and show that power dynamics shape how leaders respond to advice (12), and that when leaders listen poorly to safety concerns (17,18), junior team members (i.e., individuals with less authority) are less likely to engage in safety voice, or delay speaking-up (19), which impairs safety management. Thus, safety voice and power distance are recognised as primary causes of organisational accidents (8–10,20,21), and a range of interventions for reducing power distance in teams and enhancing speaking-up (e.g., psychological safety, training (22–24)) have been developed to improve safety voice and safety listening.

Yet, although laudable in intention, interventions to reduce power distance and increase safety voice, despite being widely advocated, have little real-world evidence from accidents. Research has not established the extent to which an absence of safety voice, or poor safety listening, have directly contributed to accidents where actors (e.g., flight crews, patients) experienced serious threats to life (e.g., 25) outside of hypothesised or simulated scenarios and isolated accident investigations (e.g., 6,26). Determining this is essential for testing the assumption that safety voice and power distance explain accident causation, and interventions that flow from this, translate to real accidents.

We address this gap in the current study, and through analysing cockpit voice recorder (CVR) transcripts of 172 historic aviation accidents, examine the role and nature of safety voice behaviours during accident scenarios. We establish to what extent safety voice i) manifests prior to accidents, ii) is ignored or dismissed by crew members, and iii) is explained by cultural norms for how junior and senior crew interact (i.e., power distance). We also consider how the introduction of CRM (24,27), an intervention designed to improve teamwork amongst safety-critical staff (e.g., flight crews, critical care teams), has increased safety voice. Our contribution is to systematically establish the role of safety voice, safety listening and power distance in the environment of real accidents, and through this, advance understanding on the extent to which these mitigate accidents.

1.1 Safety voice for safety-critical staff

Safety voice is the act of speaking-up about perceived hazards to others of equal or senior status in order to mitigate harm (1,2). Conversely, when people withhold safety concerns this is labelled ‘safety silence’ (28). The concept draws from research on communication and safety management (1) and especially employee voice research (1,2,29). This research postulates that individual team members may have critical information (e.g., on risk), and that the free flow of this information contributes to mitigating failures (30). Because of this, and the harmful consequences of poorly sharing safety information (e.g., 31,32), scholars have distinguished the concept of safety voice and provided a distinct literature (1,2,21,33,34) that extends beyond organisational environments (e.g., to non-smokers in public settings (35)), provides unique empirical data (14), relates tightly to preventing safety emergencies (in contrast to more broadranging safety related-communication during 'normal' operation (14)), and captures the communication of safety concerns that emerge from perceived risks (e.g., 13).

Safety voice is of vital importance to environments where people need to decide and act on perceived risks, such as flight crews, nuclear control room teams, critical care teams, or oil rig maintenance teams. Highlighting unsafe conditions helps to interpret the environment, create shared situational awareness (16,36), enables mitigating actions (37,38), and improves safety performance (28,39), especially when junior members of technical teams speak-up (32). For instance, in aviation, flight crews continuously handle hazardous scenarios (e.g., taking off in poor weather, addressing warning lights), and voicing and listening to concerns is deemed necessary for avoiding accidents. That is, operating aircraft requires effective coordination (e.g., to decide on risk, complete checklists, avoid opposing system input, etc. (40)) between pilots that share responsibilities for maintaining safe flight, yet have distinct tasks (e.g., flying, radio communication), information (e.g., duplicated meters may provide divergent information), experience and seniority.

Ineffective crew coordination, though rarely the sole cause, has contributed to accidents through loss of situational awareness and ineffective decision-making. For instance, status differences (e.g., strong hierarchies) and poor coordination (i.e., poor voice and listening) have contributed to fatal accidents in healthcare (e.g., the death of Elaine Bromiley after concerns about a difficult airway were dismissed (41,42)), aviation (e.g., the crash of United Airlines 173 after fuel starvation was ignored (6)) and energy (e.g., the blow out of the Deepwater Horizon oil rig after concerns about a pressure test were not raised by contractors (43)). Thus, the widespread role of communication problems in accidents underlies the growth of the safety voice literature, and the focus of interventions.

To explain why junior team members do not engage in voice, and why senior team members do not listen effectively, studies have drawn on the concept of power distance (9,23,29,44–52), which "refers to the degree to which individuals, groups, or societies accept inequalities (…) as unavoidable, legitimate, or functional" (53, p.2). Studies indicate unfavourable effects of power distance for communicating issues to leaders (54) and interventions aim to enable leaders to listen better to safety voice (e.g., support, enacting change). Yet, researching this is challenging because safety voice emerges spontaneously and its infrequent occurrence cannot be readily controlled (e.g., prompting voice could bias findings (14)). To address this, and because introducing real hazards is unethical (55), research has assessed safety voice through interviews, focus-groups and surveys (e.g., prompting memories (28,56)), vignettes (13), high-fidelity simulations (e.g., during technical procedures (15,57–59)), simulation-based training (22,24,60,61) and through laboratory experiments (e.g., presenting risks that do not require specialised technical knowledge (14)). These approaches have led to the insight that safety voice can be promoted (in terms of likelihood or onset) through leaders acting in low power distance ways. For instance, through providing encouragements, using inclusive language (15,57) or shallower hierarchies (19). Furthermore, this research indicated that risk perceptions are necessary for successful interventions (14), and that safety voice emerges after a decision on the trade-off between the benefit of mitigating harm and the cost of leaders' poor safety listening (13).

Yet, crucially, these methods assume generalisability to actual accidents, and insights on the extent to which, and how precisely, safety voice and safety listening contribute to real accidents remain scare and limited (14,19,54). Moreover, the role of power distance for safety voice during naturally occurring scenarios remains an assumption. For instance, studies have tended to use data on the occurrence of accidents instead of behaviour during accidents (8,10,62), selected a limited number of case studies (e.g., 16), or relied on inquests (26,63) that may poorly capture behaviour because self-report data reflects participants' perspectives on historic events (14). Moreover, real hazards may elicit more visceral and distinct behavioural responses than vignette and simulator studies. These limitations are consistent with meta-analyses that indicate the psychological effects established in controlled settings (e.g., simulation or laboratory studies) can be substantially different in the field and vary in their direction (64). For instance, and whilst flexible approaches to methodological realism (i.e., the extent to which methods meaningfully reflect naturally occurring scenarios) are appropriate (i.e., what makes scenarios 'real' is poorly defined and often involves an individual perception (65)), the strength of intervention effects on safety-related behaviour are stronger in archival data (capturing behaviours in the field) than self-reports (66).

This is important for theory because, whilst especially simulator evidence has provided important behavioural data (15,58), we do not know to what extent available evidence accurately represent safety voice and safety listening during true accidents, and the problem posed by the behaviours may be overestimated (e.g., if the frequency of safety voice is biased). Subtleties, like the strategies used to voice safety concerns and the ways in which voice is dismissed, have, to our knowledge, never been investigated during real accidents. This has led to the widespread assumption that a lack of safety voice is a substantial contributor to accidents, and is therefore important for mitigating declining conditions, errors and accidents (e.g., employee voice, safety voice, psychological safety (29,33,67)), and a function of wider organisational environments (e.g., safety culture, safety citizenship (20,68)). However, to date, there is no systematic exploration of the extent to which a lack of safety voice and poor listening contribute to serious accidents (19), and the level of influence exerted by power on safety voice (rather than, for example, human error) remains a proposition (e.g., 49).

Thus, whilst safety voice theory aims to explain how the behaviour contributes to accidents, and to develop interventions for improving speaking-up (69), there is a lack of data on to the extent to which, and how precisely, safety voice manifests and is listened to during real accidents. Given the conceptual importance of safety voice and safety listening as a frame for explaining failures in safety management, and for training programs aiming to improve coordination on safety (e.g., crew resource management, CRM; TeamSTEPPS (24,70)), it appears essential to consider their actual role in accident causation. For instance, without this, it is unclear how field-based behaviour should be mapped unto survey findings, or vice versa. Thus, in the current study we evaluate safety voice, safety listening and the role of power distance prior to real accidents.

1.2 The current study

Here, we investigate the extent to which safety voice varies during actual hazards that pose extreme risk, and how this is shaped by safety listening and power distance. This can be achieved through analysing transcripts from cockpit voice recorders (CVRs) from historic aviation accidents. CVRs were designed to capture and interpret sounds during accidents (e.g., flight crew communication, cues on hazards (71)), and research on flight crew communication (59,72–76) indicates CVR transcripts can be used to analyse in-situ interactions between flight crew. Our purpose is to identify behavioural patterns during aviation accidents, and normal flights and close calls are therefore out of scope. Thus, utilising CVR data, we develop exploratory hypotheses to enable an investigation into safety voice during aviation accidents and the extent to which safety voice is negatively impacted by poor safety listening and power distance.

1.2.1 Safety voice during aviation accidents

Safety voice occurs in the context of hazards, and the mitigation of risk through speaking-up is central to the concept of safety voice. Typically, hazards are studied through actual risk being hypothesised (e.g., for vignettes, simulations (13,19)) or controlled (e.g., for laboratory scenarios (14)), and eliciting risk perceptions. This revealed that stronger risk perceptions are associated with more safety voice (13,77,78). Yet, through presenting scenarios with minimal genuine risk, the extent to which the impact of risk perceptions on safety voice generalises to actual hazards remains undetermined (19). Because of this, we do not know the degree to which visceral affective risk perceptions (i.e., strong emotional responses to hazards such as dread, fear) elicit safety voice (14,79). Establishing this is important because behavioural variations can indicate when intervention may be successful (e.g., if power distance shapes safety voice). Conversely, ubiquitous or infrequent safety voice prior to accidents would suggest, respectively, that the behaviour is ineffective (i.e., because accidents occurred despite safety voice) and interventions should improve safety voice's effectiveness (e.g., when recipients listen), or that speaking-up does not pose a problem for accident causation (e.g., because it always mitigates harm, or risk simply does not elicit safety voice in practice).

We propose that actual hazards, and especially fatal accidents, should lead to more safety voice than typically established in the literature (i.e., approximately 44% of concerns are raised (1)) because cognitive evaluations of risk and visceral affective responses motivate stronger behavioural responses to mitigate harm. Probabilistic risk models highlight that hazards emerge from the accumulation of sociotechnical factors (e.g., systems design, unsafe acts (80,81)), with greater risks (i.e., impact and likelihood (82)) increasing the need for mitigating action. Yet, technical properties of risk are often difficult to evaluate (e.g., because information is ambiguous (83)) and the psychometric literature on risk perception therefore highlights that responses to hazards are rooted in analytic and affective risk perceptions (84,85). Visceral affective states emerge where encountered risks are fatal, involuntary and personally relevant, with affect heuristics providing a strong motivation to alter unsafe conditions (79,84,86). This is important, because safety voice theory often explains behaviour in terms of employee motivation (e.g., safety participation (66) or safety citizenship (87)), and little analysis has considered motivations that emerge from potentially fatal contexts. Furthermore, high costs of speaking-up may be rationally traded-off with the larger cost posed by fatalities (1,78) because the higher expected utility of speaking-up increases voice (88). Thus, and in contrast to the literature's assumption that accidents emerge from relatively low levels of safety voice (1–3,8–

[10,32]), flight crew may be expected to frequently engage in safety voice due to the extreme level of risk posed by accidents.

H1a: Flight crew engage in high levels of safety voice across historic aviation accidents.

Furthermore, we examine whether flight crew engagement in safety voice has changed over the course of history. Within the safety literature, the training of interpersonal skills is widely seen as key for improving safety voice and safety-related attitudes (89), and in aviation such training has been in place since the early 1980s through the implementation of CRM programs (24,90). Over time, these training programs became widespread (89,91) and increased in effectiveness through emphasising the design of social environments (e.g., teamworking and organisational culture) in addition to the correction of human error (90). CRM implementation may therefore be expected to have increased flight crew engagement in safety voice, and establishing this within the CVR data may inform the effectiveness of interventions for increasing safety voice.

H1b: Flight crew engagement in safety voice prior to historic aviation accidents increased since the 1980s.

1.2.2 Poor safety listening

Because safety voice is aimed at others of equal or senior status, the field has aimed to identify leadership practices favourable for speaking-up (92). Ample research indicates that when seniors listen effectively to safety voice (e.g., acknowledging and acting on concerns, versus ignoring or dismissing concerns) this promotes subsequent voice (17,18). For instance, junior team members are more likely to speak-up (18), or to do this sooner (19), when leaders are expected to listen (17,23) and indicate that speaking-up is appropriate through acting in inclusive and encouraging ways (15,57,93). However, leaders can tend to poorly listen to advice from junior team members (e.g., due to the social cost of advice-taking (12)). This suggests that even if safety voice occurs frequently it may not be listened to, with poor safety listening (i.e., ignoring or dismissing safety concerns) emerging when concerns are deemed inappropriate (e.g., when concerns are considered as factually incorrect or violating social norms (11)). For instance, no relationship between safety voice and safety listening would indicate safety voice is better predicted by risk perceptions than interpersonal dynamics. Conversely, when poor safety listening reduces safety voice this suggests that risk perceptions only partly explain safety voice and that social motivations shape the behaviour, even during extreme personal risk. If so, unique interventions are required for safety listening as a distinct contributor to accidents, and safety voice behaviour would be central to situated sense-making on risk: people share and decide on perceptions about encountered hazards, with voicing and listening to safety concerns providing two distinct aspects of a larger phenomenon capturing on-going, dynamic safety conversations. Evaluating safety listening is therefore important for conceptualising safety voice. Specifically, because poor listening may inform how leaders are expected to listen to subsequent voice (17,23), it may reduce safety voice for junior flight crew.

H2a: Safety listening increases safety voice engagement for junior team members prior to aviation accidents.

Furthermore, we also examine whether safety listening has changed over the course of history for flight crew. CRM training goals include improving how leaders engage in effective coordination on safety information inside the cockpit (24). Thus, because CRM became more widespread and effective (89–91), it may be expected that safety listening improved, and establishing this is important for enabling interventions that make safety voice more effective.

H2b: Flight crew safety listening prior to historic aviation accidents improved since the 1980s.

1.2.3 The role of power distance

Hazardous situations can provide technical and social factors contributing to risk (94,95), and safety voice may be shaped by norms that outline how juniors communicate concerns to seniors. Ample research indicates that egalitarian relationships between leaders and followers promote open communication, and whilst the operationalisation of culture through dimensions is debated (e.g., dimensions underrepresent cultural heterogeneity (96,97)), power distance has provided fruitful hypotheses to explain variation in indicators of safety performance such as accident rates (8,10), fatalities (62) and safety culture (20).

Furthermore, power distance has been considered in relationship to voice (29,46–49). For example, flat hierarchies (e.g., 98–100) and a constructive 'tone at the top' (101) promote safety voice (cf. 102), and evidence indicates that employee's power distance orientation (an individual-level construct) reduces voice (44,45,51,52). Hofstede's power distance (7) may therefore provide a valuable proxy for investigating both safety voice and safety listening on the flight deck. Yet, little behavioural evidence exists, especially during actual hazards, because research on the individual-level metric of power distance orientation (44,45,51,52) has only measured recalled and anticipated voice (i.e., which are at least one empirical step removed from behaviour (14)). In addition, studies of country-level power distance investigated the impact on the occurrence of accident instead of communication amongst flight crew (i.e., safety voice was only hypothesised as a potential explanatory variable for the occurrence of accidents (8)).

The power distance proposition for accident causation suggests that power distance explains accidents rates (8) because strong norms dictate deference to seniors' authority on safety issues (51), which are ultimately their accountability (2). This reduces safety voice for junior flight crew (9) through high power distance "(i) discouraging the correction of errors by superiors, (ii) placing primacy of communication and debate on a superior, (iii) generating unwillingness to challenge authority, and (iv) creating asymmetrical communication between management and subordinates." (20; p.775). Additionally, safety listening may explain the relationship between power distance and safety voice because violating social norms can elicit anger (11) and the perceived social cost for taking advice (e.g., appearing incompetent (12)) may be higher and elicit stronger responses to juniors speaking-up where power distance is higher. However, in the absence of direct evidence, we currently do not know the role of power distance for safety voice and safety listening during critical incidents. Thus, here we examine whether wider social norms on power distance shaped behaviour in the cockpit. In line with the power distance proposition for accident causation, we expect that power distance elicits worse safety voice and safety listening, with safety listening expected to mediate the relationship between power distance and safety voice.

H3a: Power distance reduces junior flight crew engagement in safety voice.

H3b: Power distance leads to poor safety listening.

H3c: The relationship between power distance and safety voice is mediated by safety listening.

2. METHOD

2.1 Dataset

A new dataset was generated from transcripts available in published air crash investigation reports. By January 2018, 372 transcripts were obtained from three online databases (103–105). After removing duplicate, irretrievable and non-English transcripts, the final dataset contained 172 transcripts, with a total length of 21,626 lines of transcript. All included transcripts were in English, including transcripts translated from the original recorded audio by accident investigation bodies.

The data extracted from included transcripts was: i) flight number, ii) date of incident, iii) audio source, iv) airline country registration, v) incident airspace, vi) flight phase, vii) crew and passenger numbers, viii) fatalities, ix) damage, x) attributed causal factors, xi) transcript conversational turn (i.e., all the words spoken by a speaker before another speaker starts to speak), xii) speaker. To provide interpretative context, narrative summaries and legends were included. In addition, each transcript line was coded using transcript legends and a coding scheme in terms of: i) turn number (i.e., sequential within transcripts), ii) turn type (i.e., conversation, background sounds, notes/information), iii) conversational turn number (i.e., sequential for conversation turns within transcripts), iv) role of person speaking (captain, first officer, flight engineer, flight crew with unclear role, cabin crew, air traffic control, other aircraft, ground operations, other), v) the hazard raised (i.e., if one was raised, using the words of the conversational turn), vi) how others listened to the hazard raised (action, affirmed, disaffirmed, ignored, unclear), and vii) the type of hazard based on air traffic control classification schemes (i.e., ATC interaction, Crew interaction, Distraction, Equipment/fuel, Location, Manoeuvring, Weather, Pilot actions, Planning, Company actions, Other/unclear (106))[1].

Accidents in the dataset occurred between 1962 and 2018 with 97% of the cases leading to substantial damage or the destruction of aircraft, and fatalities totalling 11,001. A crude estimation puts this at approximately 15% of historical aviation fatalities in commercial and corporate aviation since 1962[2]. Most accidents occurred on approach (32.0%) or *en route* (32.0%) and were attributed to pilot actions (32.6%), see Table 1. Flights had an average crew of 7.120 (*SD* = 5.182) and 89.701 passengers (*SD* = 97.018), with on average 42.095 survivors (*SD* = 90.191). Included flights were from airlines registered in 42 countries with an average power distance of 49.103 (*SD* = 17.043; *range*: 11-104; *skewness* = 1.157, SE = .194).

---

[1] The NATS causal factor scheme is specific to aviation incidents but may map unto typologies with a broader application. For instance, unto levels 1-3 of the Human Factors Analysis and Classification System (127): 1) unsafe acts (*Manoeuvring, Pilot actions*), 2a) unsafe environmental preconditions (*Weather, Location, Equipment/fuel*), 2b) unsafe operator preconditions (*distraction*), 2c) unsafe personnel preconditions (*ATC interaction, Crew Interaction*) and 3) unsafe supervision (*Company actions, Planning*).

[2] Aviation-Safety Network lists 66,682 historical fatalities in commercial and corporate flights between 1962-2018 (128), yet the full number of aviation fatalities is uncertain.

Table 1. Attributed causes of included accidents.

Attributed cause	n	Example
Pilot actions	56	Error during demonstration flight of Air France 296Q.
Equipment/fuel	37	Avianca 52 crashed after poorly managed fuel starvation.
Crew interaction	33	Miscommunication about arming spoilers during landing contributed to the crash of Air Canada 621.
Company actions	29	Poor CRM training provided an unfavourable environment that enabled TAM 3064 to crash due to poor coordination.
Distractions	26	Whilst distracted by a malfunction in the nose landing indication system, Eastern 401 noticed an unexpected descent too late.
Weather	26	American 1420 crashed whilst attempting to land in a thunderstorm.
ATC interaction	18	Ambiguous radio communication led Air Inter 148 to hit a mountain.
Planning	11	Poor de-icing protocols led to ingested ice, power loss and the crash of SAS 751.
Manoeuvring	7	A test flight turned into a fatal stall for Airborne Express 827.
Location	6	Texas International 655 crashed into a mountain whilst not fully using all available navigational tools.
Other/unclear	22	A bomb hit Air India 182.

n = 164 (8 missing). Total causes exceed 172 because multiple causes could be attributed.

Transcript text was based on audio sources from Cockpit Voice Recorders and/or Air Traffic Control radio communication and existed of conversational turns (n = 19,393, m = 112.750; SD = 124.829) and other data (n = 2213; m = 12.866; *SD* = 14.452; e.g., background sounds, transcriber notes). Flight crews (i.e., captains, first officers, flight engineers) provided 74.3% of the conversational turns (see Table 2). For the current study, the data was limited to conversational turns from flight crew with an identified role (i.e., conversational turns from captains, first officers, flight engineers; n = 14,128), with analyses performed on aggregated and nested data to address the hierarchical nature of the data (i.e., conversational turns within transcripts). Transcripts averaged 106.001 conversational turns (*SD* = 51.727, range: 1-641). Four transcripts had less than 5 conversational turns. The full and coded dataset is available and submitted for publication as data-in-brief.

Table 2. Frequencies of role for speakers of conversational turns.

Speaker	**n**	**Percentage**
Junior flight crew	7403	39.00%
Captain	6725	35.44%
Air traffic control	2575	13.61%
Flight crew (role unclear)	1027	5.43%
Other aircraft	476	2.52%
Other	310	1.64%
Ground operations	236	1.25%
Cabin crew	215	1.13%
Missing	471	-
Total conversational turn	19393	

2.2 Measures

Safety voice. Research assistants were trained on recognising safety voice through discussing illustrative examples and problematic cases, and the application of the coding scheme. For each conversational turn, they coded whether a hazard was raised and described the hazard in the words of the speaker. *Safety voice* (1) was coded if an individual raised a potentially dangerous situation (e.g., fire, equipment, weather, navigation, air traffic control clearances, loss of situational awareness, etc.) or indicated they were concerned. Otherwise, conversational turns were coded as *not safety voice* (0) instead of 'safety silence' (i.e., this requires data on the extent to which flight crew were concerned (14)). Standard communication procedures (e.g., going through checklists) were not coded as safety voice, unless a concern was raised. Illustrative examples are provided in Table 3 and the coding framework is submitted as Data-in-Brief. Good interrater reliability for safety voice was indicated for two randomly selected transcripts providing 291 conversational turns (*Gwet AC1* = .62, 95CI: .53-.71). The 'proportion of safety voice' was calculated as the number of conversational turns in which flight crew engaged in safety voice divided by the total number of conversational turns within a single transcript.

Safety listening. For every conversational turn containing safety voice, research assistants coded how others responded within the following three conversational turns (for illustrative examples, see Table 2). If a response to safety voice remained absent it was coded as ignored (0), if others disagreed or responded negatively it was coded as disaffirmed (-1), and favourable responses were coded as verbally affirmed (1) or immediate action (2). Indicating construct validity, low scores on safety listening were associated with accident investigation reports attributing the accident to crew communication (Spearman $r = -.156$, $p = .050$). The degree of safety listening was calculated as the average response within a single transcript.

Nested analysis ensured assumptions of independent observations were addressed (e.g., conversational turns within transcript).

Seniority. Seniority for flight crew was calculated based on the speaker of a conversational turn being senior (captain) or junior (first officer, flight engineer). Due to technical progression of aircraft, flight engineers have become less prevalent and the junior flight crew roles were therefore collapsed.

Power Distance. Power distance was operationalised through Hofstede's Power Distance Index (PDI (7)). The national background of individual pilots could not be ascertained, yet individuals' behaviour is impacted by the national culture of organisations they work for (107). Thus, PDI scores from 2015 (108) were obtained for airlines' country registration where available, bar a United Nations flight.

Table 3. Illustrative extracts from CVR transcripts for safety voice and response to safety voice.

Behaviour	Response	CVR transcript extract		
		Case	Speaker	Conversational turn
Not safety voice	n/a	Korean Air 8509	FE	*Before take-off check list complete*
			FE	*Stabilized*
			CAP	*Set take-off thrust*
			FE	*Set*
Safety voice	Disaffirmed	Surinam 764	FO*	*I think you're... according to that runway you look like you're high.*
			CAP**	*Now it's okay.*
			FO	*Slightly left of runway.*
			CAP	*Okay.*
	Ignored	Air Canada 621	FO*	*Here we have a green. The VASIS appear to be a little bit high but you are low on the glide path*
			FO	*Takes a whole airfield that way*
			CAP	*Yeah*
			CAP**	*Okay*
	Affirmed	Tower Air 41	FO*	*I don't guess you'll be able to get much of a run-up.*
			CAP**	*No. Just do the best we can. If it starts to move, we're going to take it.*
			FO	*I see an airplane looks like it's clear down the end.*
			FE	*Body gear steer?*
	Immediate action	United Airlines 173	CAP	*We can't make Troutdale*
			FO*	*We can't make anything*
			CAP**	*Okay, declare a mayday*
			FO (Radio)	*Portland tower United one seventy-three heavy Mayday we're, the engines are flaming out, we're going down, we're not going to be able to make the airport*

* Conversational turn containing safety voice. ** Key message for the response. CAP: Captain, FO: First Officer, FE: Flight Engineer.

3. RESULTS

3.1 Safety voice during aviation accidents

Supporting hypothesis 1a, flight crew safety voice was near ubiquitous across accidents, but the proportion of safety voice varied within transcripts. Safety voice occurred in all but two of the accidents (95CI: 97.2-100.5%), with only two accidents having no instances of safety voice (i.e., Air India 182, TAM 3054). This was not statistically different from 100% ($t(170) = -1.418$, $p = .158$). The proportion safety comprised, on average, 14.19% of the transcripts (95CI: 11.79-16.59%; $t(170) = 11.668$, $p < .001$). The proportion safety voice for flights where someone spoke-up, and that contained more than five conversational turns, ranged from 1.13% (Asiana Airlines 214) to 67.3% (PSA 182).

The proportion of safety voice was not predicted by attributed accident causes (Wilk's Lambdas $\leq .999$, $Fs(2,139) \leq 1.000$, $ps \geq .379$, $\eta^2 \leq .014$), and did not alter the extent of damage the plane incurred ($F(4,50) = 1.562$, $p = .199$, $\eta^2 = .111$). Yet, the degree to which flight crew engaged in safety voice changed over time, but surprisingly rejecting hypothesis 1b, the degree of safety voice became *less* overall ($b = -.007$, $F(1,166) = 55.812$, $p < .001$, $R^2 = .252$), see Figure 1. This trend was consistent with accidents over time being more frequently attributed to poor crew interaction ($OR = 1.065$, $Wald(1) = 9.387$, $p = .002$). Flight crew were less likely to engage in safety voice during historic accidents after the introduction of CRM in approximately 1981[3] ($F(1,166) = 56.260$, $p < .001$, $\eta^2 = .253$).

---

[3] The year 1981 was chosen because CRM programs emerged in the early 1980s (24). Yet, it should be noted that CRM was not simultaneously introduced across airlines.

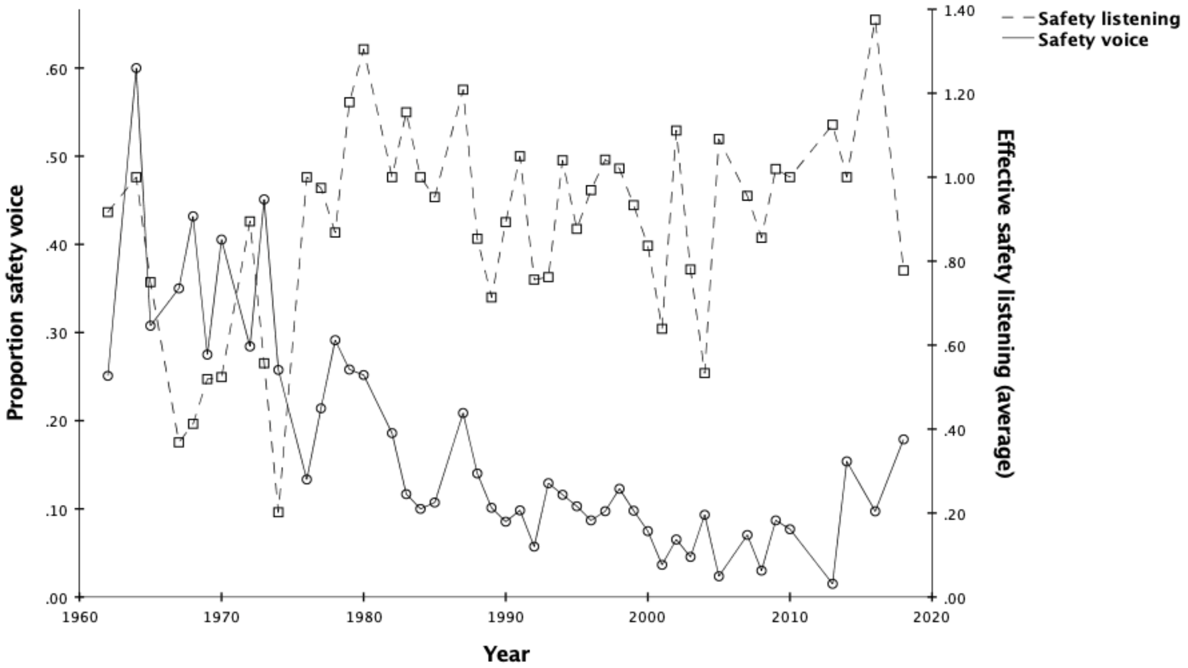

Figure 1. Historic trends of the proportion of safety voice and average response to safety voice within CVR transcripts.

Given the near obliquitous occurrence of safety voice acts identified, we describe cases to illustrate that the effectiveness of voice depends upon technical issues and safety listening. Often, crews voiced concerns too late. For instance, USAF 27, where the co-pilot mentioned the potential for bird strike by saying “lot of birds here”. The captain acknowledged (“Lotta birds here”), however the crew did not respond quickly enough to the hazard, leading to 24 fatalities. Conversely, for SAS 751, the first officer voiced several times during an ongoing event (e.g., "We have problems with our engines, please... we need to go back to, ... to go back to Arlanda"), and despite the crew recognising the problem they could not resolve it because the problem was an underlying technical issue (ice on wings, with engine ice ingestion). Finally, for Saudia 163, safety voice was repeatedly engaged in (e.g., by the first officer continually raising concerns about smoke in the cabin). The captain and crew responded to this, however poor coordination amongst the crew contributed to the accident.

3.2 Poor safety listening

Supporting hypothesis 2a, poor safety listening reduced the overall proportion of safety voice in a transcript ($\beta = -.200$, $F(1,156) = 6.499$, $p = .012$, $R^2 = .040$) and specifically for junior flight crew speaking-up ($\beta = -.212$, $F(1,109) = 5.105$, $p = .026$, $R^2 = .045$). Listening behaviours (n = 1090) tended to be favourable but varied across accidents ($M = .890$; $SE = .030$; $t(157) = 29.218$, $p < .001$): 82 accidents (e.g., Alaska airlines 261) only saw effective safety listening, 3 only one negative response (i.e., Aviation services, Crossair 498, Martinair 492), and 33 accidents saw repeated poor listening (range: 2-33 times; e.g., Texas International 655). Junior flight crew were listened to less, compared to senior flight crew ($F(1,229) = 1.345$, $p = .002$, $\eta^2 = .264$).

The degree of safety listening was not predicted by attributed accident causes (Wilk's Lambdas ≤ .999, $Fs(2,139) \leq 1.000$, $ps \geq .379$, $\eta^2 \leq .014$), yet poor listening led to more plane damage ($b = .359$, $F(1,151) = 8.697$, $p < .001$, $R^2 = .054$). Moreover, and supporting hypothesis 2b, safety listening became more favourable over time ($F(1,133) = 1.685$, $p < .001$, $\eta^2 = .191$), with the introduction of CRM providing a strong historic turning point because listening became more favourable on average after this ($F(1,133) = 1.685$, $p < .001$, $\eta^2 = .191$), see Figure 1.

To illustrate the nature of safety listening, we report on examples of poor listening. For instance, for Kalitta 808 (crashed after stalling), two warnings by a flight engineer about low airspeed ("You know, we're not getting' our airspeed back there" and "Watch the, keep your airspeed up") were ignored by the crew, who were focussed on identifying the strobe light for landing (e.g., in response to concerns the captain asked "Where's the strobe?"). Similarly, for TWA 514 (crashed due to flying at an unsafe altitude), repeated attempts by the first officer to share concerns about the altimeter ("I hate the altitude jumping around"; "Gives you a headache after a while, watching this jumping around like that") were not acknowledged by the captain,

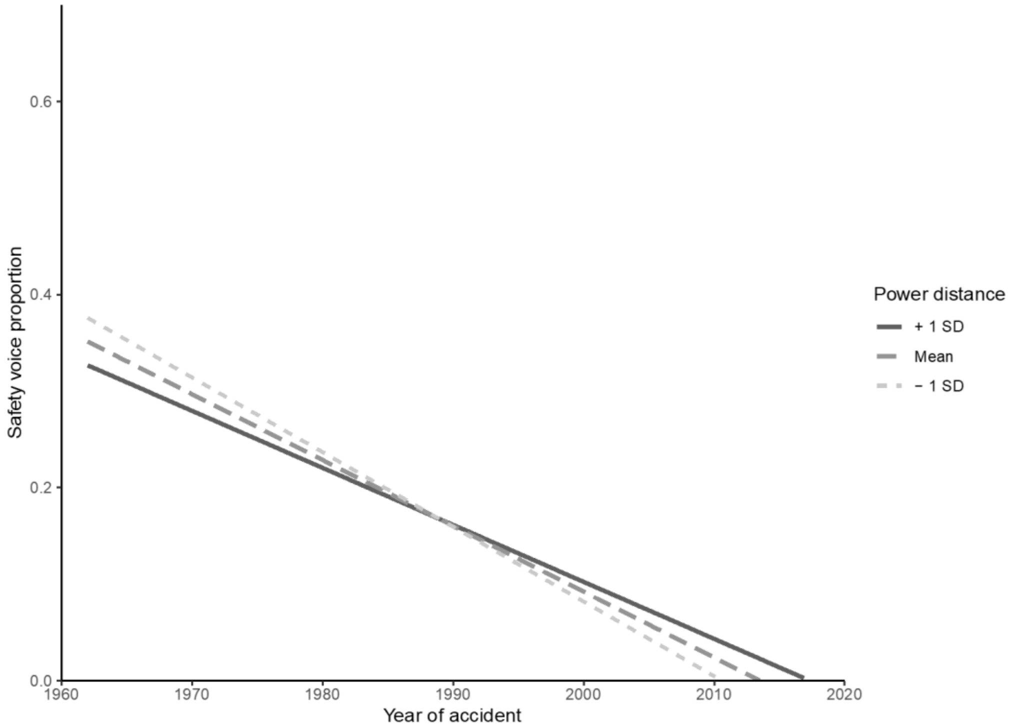

Figure 2. The proportion of safety voice within a transcript given the year of the accident and airline power distance.

who was focussed on visually identifying the ground. In other cases, safety voice led to disagreement: during landing in a Metro II aircraft the first officer voiced on the landing gear "is it down?", which led to an unresolved confusion between the captain ("yeah gear's down") and the co-pilot ("No its up"). Similarly, for Aeroflot 9981, a co-pilot's request to disengage from a dangerous landing ("No, let's...go around") was first dismissed by the captain ("Why are we going around?") and then confirmed too late ("Tell them "go around").

3.3 The role of power distance

Power distance increased the likelihood that accidents were attributed to crew interaction (*OR* = 1.031, *Wald*(1) = 7.856, *p* = .005), but surprisingly power distance only explained the extent of safety voice (supporting hypothesis 3a), not safety listening (rejecting hypothesis 3b). The proportion of safety voice in a transcript was not predicted by direct linear effects for the seniority of the voicer (*OR* = 1.051, *Wald*(1) = .720, *p* = .396) and power distance (*OR* = 1.00, *Wald*(1) = .002, *p* = .964), and as shown in Table 3, this emerged due to an interaction-effect between seniority and power distance on safety voice (*OR* = 1.003, *Wald*(1) = 4.302, *p* = .032).

Indicating that norms for engaging with seniors shape safety voice, power distance predicted safety voice (proportion of safety voice in a transcript = -.118 + .0212(PDI) – $4.740*10^{-4}$(PDI^2)+ $3.123*10^{-6}$(PDI^3), $F(3,151) = 3.104$, $p < .001$.), and predicted more safety voice for junior flight crew in low power distance countries ($OR = .992$, $Wald(1) = 4.487$, $p = .034$), whereas senior flight crew voiced more in high power distance countries ($OR = 1.006$, $Wald(1) = 4.397$, $p = .036$). To illustrate this interaction: junior flight crew were 1.494 times less likely to engage in safety voice with a 50-point increase in power distance (i.e., half the scale). Moreover, as illustrated in Figure 2, the identified historic decline in the extent of safety voice was especially strong for low power distance countries: a strong interaction-effect existed for power distance and year on the proportion of safety voice in a transcript ($F(34,50) = 3.262$, $p < .001$, $\eta^2 = .689$).

Surprisingly, power distance was not associated with poor safety listening to junior flight crew speaking-up ($r = -.041, p = .681$), with only a weak association (Spearman's $r = -.071$, $p = .033$) indicating that voice may have been less ignored in high power distance airlines because it involved a more extreme act (providing minimal support for hypothesis 3b). Furthermore, and rejecting hypothesis 3c, safety listening did not explain the effect of power distance on safety voice because no mediation-effect was found in general ($b = .000$, $SE = .002$, *95CI:* -.004 – .005) or for junior flight crew specifically ($b = .008$, $SE = .025$, *95CI:* -.028 – .071), and no interaction-effects existed for power distance with seniority on safety listening ($F(83,703) = .989$, $p = .510$, $\eta^2 = .105$) and with safety listening on the proportion of safety voice in a transcript ($F(1,141) = .540$, $p = .464$, $\eta^2 = .004$). However, indicating a moderation effect of power distance and consistent with the reduction in safety voice, an interaction-effect indicated that safety listening became more favourable over time for low power distance airlines ($F(22,829) = 2.057$, $p = .003$, $\eta^2 = .052$).

4. DISCUSSION

Through providing the first systematic and behavioural analysis of safety voice prior to aviation accidents, we tested the prediction that high levels of risk lead to more safety voice during actual historic aviation accidents, with effective safety listening and the introduction of CRM training improving the extent to which junior flight crew engaged in safety voice. Furthermore, we provided the first behavioural examination of the power distance proposition for accident causation suggesting that power distance reduces safety voice through less safety listening. In support of these predictions, we demonstrated that safety-critical staff nearly always speak-up across hazardous situations. Initial acts of safety voice within a transcript were frequently listened to poorly, and this reduced the amount of subsequent safety voice prior to accidents. Power distance explained the extent of safety voice, yet no direct linear or mediation effects were found for safety listening. Moreover, the introduction of CRM training only led to changes in annual trends on safety voice and safety listening where power distance was low. These findings have important implications for safety voice theory and safety management.

4.1 Theoretical implications

We provided the first evidence that people engage in *real* safety voice behaviour during *genuine* accidents, and indicated they do this nearly always across accidents. This is important because through relying on data from simulated or imagined scenarios the safety voice literature has assumed that accidents can emerge from a lack of safety voice (1,2,32), yet we indicated that accidents still occurred despite flight crew speaking-up. Thus, in contrast to prevailing thought, we indicate that accidents cannot be assumed to necessarily emerge from a lack of safety voice, or that the behaviour is sufficient for avoiding harm. This means that through relying on selective case studies, inquests, simulations and studies operationalising hazards (8,10,16,26,62,63), research has provided insufficient insights on behaviour in the field and wrongly assumed the central problem is a simple absence of safety voice. Thus we contribute a fundamental insight that research should be grounded in the analysis of safety voice during actual hazards, and progress from making safety voice more likely to making safety voice more effective (i.e., for preventing harmful outcomes (3,32)).

Most importantly, we indicated that safety concerns were often ignored or rejected, and this suggests that safety listening may be conceptualised as an essential step in the chain between hazards eliciting concerns, people raising concerns and threats being mitigated (see Figure 3). Whilst the safety voice literature has previously established that anticipated responses from leaders are important (15,17–19,23,57,93), explicitly conceptualising safety listening is important because its role in mitigating safety threats and making safety voice more effective has remained underdeveloped. Part of listening effectively to safety voice is responding in constructive ways (e.g., taking action, demonstrating personal interest (110)), which may confirm risk perceptions and enables more voice (111). We support the generalisation of research on leaders' poor safety listening from controlled environments (e.g., 15,57) through demonstrating that the degree to which flight crew spoke-up during aviation accidents was lower when previous concerns were poorly listened to. Thus, we indicate the need for novel interventions on safety listening and enable the application of concepts such as psychological safety (67) and advice taking (12) to real accidents, and we suggest future research investigates how safety voice

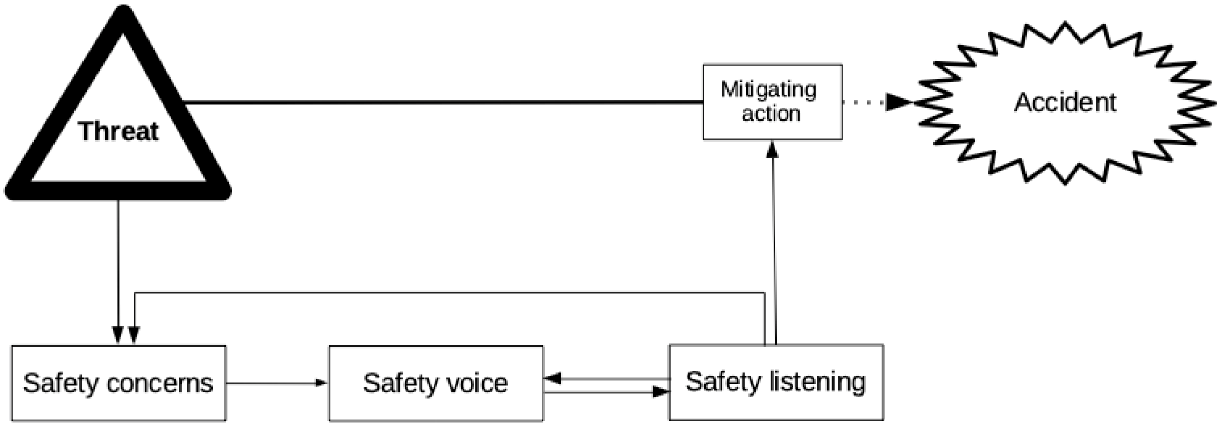

Figure 3. Threat Mitigation model of safety voice.
Note: the model highlights that the dysfunctional momentum of threats towards accidents (37) can be mitigated (dotted line), when threats elicit higher degrees of safety concerns, safety voice and safety listening.

can be made more effective through distinguishing between safety voice and safety listening, and the design of interventions that enable recipients to enact change (112). For instance, through exploring how the concept of loss aversion (113) explains effective listening when people perceive risk.

The near ubiquitous occurrence of safety voice across accidents is consistent with the notion that the perception of risk provides motivation for sharing situational awareness and initiating decision-making (37,66,79,84), and supports vignette-based and experimental findings indicating that risk is central to safety voice (13,14). However, few safety voice studies have assessed risk or delineated leading indicators of accidents (e.g., unsafe acts, or their preconditions (94,95)). Yet, because we indicated that safety voice is more prevalent during accident than typically established in the literature[4], this indicates a need for outlining how findings from other methodologies (e.g., surveys, interviews, experiments (1)) may be mapped unto real hazards in terms of distinct sociotechnical risk factors (114). For instance, future studies may enable the comparison of safety voice and safety listening across hazardous situations through carefully describing how hazard characteristics (i.e., in terms of technical or physical properties and levels of risk) elicit visceral states which are difficult to recall or forecast (14,79).

In addition, we found that whilst safety voice occurred across accidents, the amount of safety voice varied across transcripts. This is important for safety voice theory because it confirms that factors beyond physical risk influence the degree to which people speak-up about safety (1) and thus, whilst it is essential to increase the effectiveness of the behaviour, scope remains for increasing the degree to which people speak-up. In particular, whilst leader behaviours (e.g., power distance, leadership styles) have been proposed to cause accidents through reducing safety voice (9,29,46–49), we provided the first direct and systematic evidence that social structures

---

[4] A one-sample T-test revealed that the average proportion of safety voice in the transcripts was different from previous research that indicates that people only raise their concerns in approximately 44% of the cases ($t(170) = 66.494$, $p < .001$).

can reduce safety voice during actual accidents. Additionally, we provide an important nuance to the power distance proposition for accident causation through highlighting that power distance reduces safety voice, but not through leaders listening more poorly in high distance environments. Thus, we confirm research indicating that power distance contributes to accident rates (8,10), evidence the generalisability of findings on the individual level construct of power distance orientation (51) to influences on safety voice within safety-critical teams, and indicate the need to investigate how power distance and safety listening independently reduce safety voice. Moreover, we enabled research to investigate established safety voice antecedents (e.g., leaders using inclusive language (57)) and interventions (e.g., education-based training (69)) during real-life hazards. Finally, this contributes to the wider safety management literature (e.g., risk perception, safety citizenship, safety culture (82,84,115)) through indicating that the investigation of relatively stable latent risks (e.g., organisational culture) may be supplemented by the investigation of safety voice and safety listening because it provides access to social mechanisms explaining how people communicate during emergencies. In particular, based on this future research may identify social mechanisms that distinguish safety-related communication during ‘normal operations’ (i.e., when it is business-as-usual) from safety voice.

Fourth, we indicated that the introduction of CRM provided a good explanation of historic trends in safety voice and safety listening, yet rejecting hypotheses 1b, we found that safety voice declined over time. This is surprising because it contradicts the literature that suggests CRM improves speaking-up (24), but it may be explained by CRM improving safety listening (e.g., through increases psychological safety (17)) and thus reducing the need for repeated safety voice (i.e., because cooperative relationships increase shared situational awareness (16,36)) or even preventing accidents (and thus the inclusion in the dataset). This would support the use of CRM training, and through providing the first evidence on reduced effectiveness of CRM in higher power distance contexts, we indicated a need for research to improve CRM training across cultural contexts.

Finally, investigating safety voice and safety listening through a cultural lens can extend safety voice theory through the identification of additional cultural predictors of safety voice. Safety voice research has rarely done this (1), but this would be valuable for the design of new interventions. Future research may identify cross-cultural differences in safety voice due to face-saving (116), global differences in leadership values and practices (117), or other national culture dimensions (e.g., individualism, uncertainty avoidance, masculinity, long-term orientation (7)). Yet, it may prove more optimal to develop the concept of safety voice as an integral activity to organisational politics (118) and sense-making on risk (119,120). These approaches describe how cultural processes emerge in response to challenges for dealing with risk, and adopting them may extend existing perspectives (e.g., highlighting that safety voice results from voice climate (121)) through indicating how safety voice and safety listening dynamically constitute safety culture. For instance, through longitudinal investigations on the sense-making process through which the behaviours lead to institutional change.

4.2 Practical implications

Our results have practical implication for safety management and safety-critical teams. First, unlike previously assumed (1,2,32), safety voice occurs during accidents but its effectiveness for avoiding harm needs to improve: we indicated a gap between safety voice and the mitigation of harm. This means that whilst safety voice is necessary for avoiding accidents, it provides incomplete protection (e.g., in terms of Reason’s Swiss Cheese model (94)) without practitioners recognising and responding appropriately to concerns raised (e.g., through engaging in open conversation, taking action). Thus, whilst safety voice contributes to the mitigation of risk, steps need to be evaluated for increasing the effectiveness of safety voice, for instance through improving safety listening, and this should be incorporated into training programs such as CRM (24).

Second, our findings support the scope and benefit of CRM training programs. This is because variation in the degree of safety voice during accidents indicates interventions may improve the behaviour, and the historic introduction of CRM led to better safety listening and, as argued above, safety voice. However, whilst research has indicated the impact of cultural norms on safety behaviours and accidents (10,20,122), we indicated that CRM training remains insufficiently tailored to high power distance environments. This is especially pressing for safety management in these environments because research indicates that accidents are more likely where norms do not support egalitarian interactions (8). After research increasing CRM's overall effectiveness (27), the next phase of CRM implementation should therefore tailor training programs to specific environments.

4.3 Limitations

Five limitations exist for the current study. Below we suggest how these may be addressed and indicate steps for future research utilising the CVR dataset.

First, the analyses only enable tentative conclusions on the occurrence of safety silence and outcomes prior to accidents. Because data on normal flights and close calls is carefully protected by airlines (i.e., due to commercial sensitivity, data protection regulation), this data was not included in the dataset and this means that conclusions are not straightforward on the extent to which safety voice would have avoided harm or appeared differently during close-calls (i.e., this requires data on the relationship between safety voice and the occurrence of accidents vs close calls). Because of this the attribution of blame is not only undesirable, but invalid, and generalisations to 'normal' flight conditions should not be readily made. Safety voice theory may advance through establishing how safety voice enables the avoidance of harm, and this may be optimally achieved through triangulating the CVR dataset with data on close calls and safety performance (123). Commercial airlines may support these aims through making this data publicly available for research and safety management purposes.

Additionally, we demonstrated that safety voice occurred across accidents and because people speak-up in response to perceived hazards (14) it is highly probable that flight crew spoke-up because they perceived risk during the accidents. However, whilst conclusions on the extent of safety voice were possible, the absence of safety voice does not readily constitute safety silence (i.e., flight crew may not speak-up because they are not concerned (14)). Future research may investigate text-based measures for assessing safety concerns in flight crew speech and apply these to establish conclusions on the degree of safety silence. For instance, in future research we aim to establish the extent to which safety silence can be scaled based the degree to which people engage in safety voice. Other scholars may utilise the CVR data to investigate the impact of the assertiveness of safety voice (124) on effective safety listening through creating a more compelling need to change the dysfunctional momentum of hazards towards harm (37).

Second, the quality of the dataset is dependent on included CVR transcripts, the condition of the source files after accidents occurred, and the standard of transcription (74). Included transcripts were available at the online databases and written in English, and other transcripts may have been missed. However, the dataset incorporated approximately 15% of commercial and corporate aviation fatalities since 1962 and thus the data provides substantial coverage of known cases. Not all original audio files were accessible, and we needed to assume reasonable transcription accuracy (e.g., accuracy of words uttered, translation to English). We suggest this is appropriate because providing accurate transcripts is in the interest of accident investigations, and transcription uncertainties were indicated in the transcripts (e.g., ‘unintelligible’). Future research may enhance the dataset through extending the number of transcripts (e.g., new accidents, or from alternative sources), or directly testing the transcription quality.

Third, safety behaviours during hazardous situations are complex phenomena, and additional variables may therefore impact on safety voice and safety listening. We focused on safety voice, safety listening and power distance, but more variables should be considered. This is consistent with a recent systematic literature review indicating in excess of 32 higher-order safety voice antecedents (1) and studies indicating that safety listening is impacted by factors such as cognitive tunnelling (i.e., fixation of attention due to high workload, stress (125)). Future research may therefore add to understanding the complexities of safety voice and safety listening during hazardous situations through investigating alternative mechanisms or theoretical propositions.

Fourth, we established good interrater reliability for safety voice, yet this was based on a small subset of the data and interrater reliability may be different for the complete dataset. We aimed to provide consistent coding through employing research assistants highly familiar with observing safety voice and providing substantive training on the CVR data, and provided the CVR dataset for future research.

Finally, the appropriateness of using Hofstede's dimensions has been debated (96,97). People within countries display a broad range of psychological tendencies (126), and whilst cultures remain relatively stable, 172 accidents may not reflect the heterogeneity of cultures across and within countries. In addition, flight crews increasingly contain expats and the national culture associated with the airline may therefore not accurately capture the nationalities of individual crewmembers. Individuals' nationalities could not be ascertained, and we suggested that national-level data may be used as a proxy for power distance on the flight deck. Arguably, a measurement error due to cultural variations may suggest that the power distance effect is stronger than we established (i.e., because measurement errors could reduce power). Through presenting variation in the degree to which people raise concerns across 14,128 conversational turns from airlines from 42 countries we provide a first step in this direction. The literature may further reduce potential biases from homogenous samples through replicating these findings for other hazards and industries.

4.4 Conclusion

Safety voice is theorised as an important mitigating factor for maintaining safety, with power distance proposed to inhibit people from speaking-up. Yet, behavioural research during actual hazards has been scant. We showed that safety voice was near ubiquitous across historic accidents that posed fatal risk, whilst variation existed in the degree to which safety voice dominated conversations. This underscores the role of risk perception as a trigger for safety voice, and indicates that the literature can no longer assume that safety voice is sufficient for avoiding harm or that the behaviour is absent during accidents.

Variation in safety voice indicated the importance of contextual variables for shaping safety voice, and we demonstrated these include safety listening, power distance and the provision of CRM training. Safety voice by junior flight crew was often ignored or rejected, indicating the need for the literature to conceptualise safety listening as an essential step for the effective mitigation of safety threats. We provided the first behavioural evidence supporting the power distance proposition for accident causation through indicating higher power distance inhibits safety voice behaviour, yet this was not through poor safety listening and a need exists to establish the mechanism through which power distance reduces safety voice. Finally, hinting at the importance of CRM training for mitigating hazards, safety voice improved after the introduction of CRM training. Yet, this was only the case for low power distance countries, indicating a need for tailoring CRM training programs to high power distance environments. Across sociocultural contexts, people mitigate hazards through engaging in conversation with others, and the field needs to incorporate how people enact safety voice because raising and listening to safety concerns provide unique challenges for avoiding accidents.

ACKNOWLEDGEMENTS

This work was made possible by a PhD scholarship for Mark Noort from the Economics and Social Science Research Council (ESRC). Ethical approval was obtained from LSE's research ethics committee (#1051). The authors would like to express their gratitude for data collection and coding to Lindsie Arthur-Hulme, Nanne Houtsma, Alex Goddard, Jason Murcko and Celestin Okoroji.

REFERENCES

1. Noort MC, Reader TW, Gillespie A. Speaking up to prevent harm: A systematic review of the safety voice literature. Safety Science, 2019; 117(May):375–387.
2. Tucker S, Chmiel N, Turner N, Hershcovis MS, Stride CB, Sandy Hershcovis M, Stride CB. Perceived organizational support for safety and employee safety voice: The mediating role of coworker support for safety. Journal of Occupational Health Psychology, 2008; 13(4):319–330.
3. Bienefeld N, Grote G. Silence that may kill: When aircrew members don't speak up and why. Aviation Psychology and Applied Human Factors, 2012; 2(1):1–10.
4. Cocklin JT. Swissair 111 Human Factors: Checklist and Cockpit Communications. Journal of Air Transportation, 2004; 9(3):19–42.
5. Cookson S. ' We Need Priority Please ' Mitigated Speech in the Crash of Avianca Flight 052. 2015:578–583.
6. NTSB. Aircraft Accident Report: United Airlines, Inc. McDonnell Douglas, Dc-8-61, N8082U., 1978.
7. Hofstede G, Hofstede GJ, Minkov M. Cultures and Organizations: Software of the Mind, Revised and Expanded 3rd Ed. McGraw-Hill, New York, NY, 2010.
8. Enomoto CE, Geisler KR. Culture and plane crashes: A cross-country test of the gladwell hypothesis. Economics and Sociology, 2017; 10(3):281–293.
9. Gladwell M. Outliers: The Story of Success. New York, Boston, London: Little, Brown and Company, 2008.
10. Soeters JL, Boer PC. Culture and Flight Safety in Military Aviation. The International Journal of Aviation

Psychology, 2000; 10(2):111–133.

11. Kam CC-S, Bond MH. Emotional reactions of anger and shame to the norm violation characterizing episodes of interpersonal harm. British Journal of Social Psychology, 2009; 48(2):203–219.

12. Tost LP, Gino F, Larrick RP. Power, competitiveness, and advice taking: Why the powerful don't listen. Organizational Behavior and Human Decision Processes, 2012; 117(1):53–65.

13. Schwappach DLB, Gehring K. Silence That Can Be Dangerous: A Vignette Study to Assess Healthcare Professionals' Likelihood of Speaking up about Safety Concerns. Spilsbury K (ed). PLoS ONE, 2014; 9(8):e104720.

14. Noort MC, Reader TW, Gillespie A. Walking the Plank: An Experimental Paradigm to Investigate Safety Voice. Frontiers in Psychology, 2019; 10(April):1–18.

15. Barzallo Salazar MJ, Minkoff H, Bayya J, Gillett B, Onoriode H, Weedon J, Altshuler L, Fisher N. Influence of surgeon behavior on trainee willingness to speak up: A randomized controlled trial. Journal of the American College of Surgeons, 2014; 219(5):1001–1007.

16. Driscoll G. Cockpit conversation: A communication analysis of three aviation accidents. 2002.

17. Edmondson AC. Psychological Safety and Learning Behavior in Work Teams. Administrative Science Quarterly, 1999; 44(2):350–383.

18. Nembhard IM, Edmondson AC. Making it safe: the effects of leader inclusiveness and professional status on psychological safety and improvement efforts in health care teams. Journal of Organizational Behavior, 2006; 27(7):941–966.

19. Krenz H, Burtscher MJ, Grande B, Kolbe M. Nurses' voice: the role of hierarchy and leadership. Leadership in Health Services, 2020; 33(1):12–26.

20. Reader TW, Noort MC, Kirwan B, Shorrock S, Kirwan B. Safety sans frontières: An international safety culture model. Risk analysis, 2015; 35(5):770–789.

21. Conchie SM, Taylor PJ, Donald IJ. Promoting safety voice with safety-specific transformational leadership: The mediating role of two dimensions of trust. Journal of Occupational Health Psychology, 2012; 17(1):105–115.

22. Kolbe M, Weiss M, Grote G, Knauth A, Dambach M, Spahn DR, Grande B. TeamGAINS: A tool for structured debriefings for simulation-based team trainings. BMJ Quality and Safety, 2013; 22(7):541–553.

23. Newman A, Donohue R, Eva N. Psychological safety: A systematic review of the literature. Human

Resource Management Review, 2017; 27(3):521–535.

24. Kanki B, Anca J, Chidester T. Crew Resource Management. 3rd ed. London: Elsevier Inc., 2019.

25. Raemer DB, Kolbe M, Minehart RD, Rudolph JW, Pian-Smith MCM. Improving Anesthesiologists' Ability to Speak Up in the Operating Room. Academic Medicine, 2016; 91(4):530–539.

26. Francis R. Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry Executive Summary Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry: Executive Report. London, UK, 2013.

27. Helmreich RL, Chidester T, Foushee HC, Gregorich S, Wilhelm JA. How effective is cockpit resource management training? Exploring issues in evaluating the impact of programs to enhance crew coordination. Flight safety digest, 1990; 9(5):1–17.

28. Manapragada A, Bruk-Lee V. Staying silent about safety issues: Conceptualizing and measuring safety silence motives. Accident Analysis & Prevention, 2016; 91(November):144–156.

29. Morrison EW. Employee Voice and Silence. Annual Review of Organizational Psychology and Organizational Behavior, 2014; 1(1):173–197.

30. Westrum R. The study of information flow: A personal journey. Safety Science, 2014; 67(November):58–63.

31. Novak A. Improving safety through speaking up: An ethical and financial imperative. Journal of Healthcare Risk Management, 2019; 39(1):19–27.

32. Kolbe M, Burtscher MJ, Wacker J, Grande B, Nohynkova R, Manser T, Spahn DR, Grote G. Speaking up is related to better team performance in simulated anesthesia inductions: An observational study. Anesthesia and Analgesia, 2012; 115(5):1099–1108.

33. Okuyama A, Wagner C, Bijnen B. Speaking up for patient safety by hospital-based health care professionals: a literature review. BMC Health Services Research, 2014; 14(1):61.

34. Morrow KJ, Gustavson AM, Jones J. Speaking up behaviours (safety voices) of healthcare workers: A metasynthesis of qualitative research studies. International Journal of Nursing Studies, 2016; 64(1):42–51.

35. Bigman CA, Mello S, Sanders-Jackson A, Tan ASL. Speaking up about Lighting up in Public: Examining Psychosocial Correlates of Smoking and Vaping Assertive Communication Intentions among U.S. Adults. Health Communication, 2019; 34(4):500–510.

36. Foushee HC. Dyads and triads at 35,000 feet: Factors affecting group process and aircrew performance.

American Psychologist, 1984; 39(8):885–893.

37. Barton MA, Sutcliffe KM. Overcoming dysfunctional momentum: Organizational safety as a social achievement. Human Relations, 2009; 62(9):1327–1356.

38. Sexton JB, Helmreich RL. Using language in the cockpit: Relationships with workload and performance. Dietrich R, Von Meltzer T (eds). Communication in high risk environments, 2003; 12:57–74.

39. Manias E. Communication relating to family members' involvement and understandings about patients' medication management in hospital. Health Expectations, 2015; 18(5):850–866.

40. Skogstad, Hellesoy. Cockpit-cabin crew interaction: satisfaction with communication and information exchange. Aviation Space and Environmental Medicine, 1995; 66(9):841–848.

41. Fioratou E, Flin R, Glavin R, Patey R. Beyond monitoring: Distributed situation awareness in anaesthesia. British Journal of Anaesthesia, 2010; 105(1):83–90.

42. Bromiley M, Mitchell L. Would you Speak up if the Consultant Got it Wrong? …And Would you Listen if Someone Said You'd Got it Wrong? Journal of Perioperative Practice, 2009; 19(10):326–329.

43. Reader TW, O'Connor P. The Deepwater Horizon explosion: non-technical skills, safety culture, and system complexity. Journal of Risk Research, 2014; 17(3):405–424.

44. Huang X, Van de Vliert E, Van der Vegt G. Breaking the Silence Culture: Stimulation of Participation and Employee Opinion Withholding Cross-nationally. Management and Organization Review, 2005; 1(3):459–482.

45. Landau J. To speak or not to speak: Predictors of voice propensity. Journal of Organizational Culture, Communications and Conflict, 2009; 13(1):35–54.

46. Liang J, Farh CIC, Farh J-L. Psychological Antecedents of Promotive and Prohibitive Voice: A Two-Wave Examination. Academy of Management Journal, 2012; 55(1):71–92.

47. Wilkinson A, Barry M, Morrison E. Toward an integration of research on employee voice. Human Resource Management Review, 2020; 30(1):100677.

48. Morrison EW. Employee Voice Behavior: Integration and Directions for Future Research. The Academy of Management Annals, 2011; 5(1):373–412.

49. Kwon B, Farndale E, Park JG. Employee voice and work engagement: Macro, meso, and micro-level drivers of convergence? Human Resource Management Review, 2016; 26(4):327–337.

50. MacNab B, Brislin R, Worthley R, Galperin BL, Jenner S, Lituchy TR, MacLean J, Aguilera GM, Ravlin E, Tiessen JH, Bess D, Turcotte MF. Culture and ethics management: Whistle-blowing and internal

reporting within a NAFTA country context. International Journal of Cross Cultural Management, 2007; 7(1):5–28.

51. Botero IC, Van Dyne L. Employee Voice Behavior: Interactive Effects of LMX and Power Distance in the United States and Colombia. Management Communication Quarterly, 2009; 23(1):84–104.

52. Rhee J, Dedahanov A, Lee D. Relationships among power distance, collectivism, punishment, and acquiescent, defensive, or prosocial silence. Social Behavior and Personality, 2014; 42(5):705–720.

53. Daniels MA, Greguras GJ. Exploring the Nature of Power Distance: Implications for Micro- and Macro-Level Theories, Processes, and Outcomes. Journal of Management, 2014; 40(5):1202–1229.

54. Peadon RR, Hurley J, Hutchinson M. Hierarchy and medical error : Speaking up when witnessing an error. Safety Science, 2020; 125(January 2019):104648.

55. APA. Ethical Principles of Psychologists and Code of Conduct. Washington, DC, 2017.

56. Alingh CW, Van Wijngaarden JDH, Van De Voorde K, Paauwe J, Huijsman R. Speaking up about patient safety concerns: The influence of safety management approaches and climate on nurses' willingness to speak up. BMJ Quality and Safety, 2019; 28(1):39–48.

57. Weiss M, Kolbe M, Grote G, Spahn DR, Grande B. We can do it! Inclusive leader language promotes voice behavior in multi-professional teams. The Leadership Quarterly, 2018; 29(3):389–402.

58. Smith HPR. A simulator study of the interaction of pilot workload with errors, vigilance, and decisions. Nasa Technical Memorandum, 1979; No. 78482(January):57.

59. Foushee HC, Manos KL. Information transfer within the cockpit: Problems in intracockpit communications. P. in Information Transfer Problems in the Aviation System (NASA TP-1875). Moffett Field, CA: NASA-Ames Research Center, 1981.

60. Kines P, Andersen LPSS, Spangenberg S, Mikkelsen KL, Dyreborg J, Zohar D. Improving construction site safety through leader-based verbal safety communication. Journal of Safety Research, 2010; 41(5):399–406.

61. Leonard M, Graham S, Bonacum D. The human factor: the critical importance of effective teamwork and communication in providing safe care. Quality and Safety in Health Care, 2004; 13(SUPPL. 1):85–90.

62. Anicich EM, Swaab RI, Galinsky AD. Hierarchical cultural values predict success and mortality in high-stakes teams. PNAS, 2015; 112(5):1338–1343.

63. Francis R. Freedom to Speak up: An Independent Review into Creating an Open and Honest Reporting Culture in the NHS., 2015.

64. Mitchell G. Revisiting Truth or Triviality: The External Validity of Research in the Psychological Laboratory. Perspectives on Psychological Science, 2012; 7(2):109–117.

65. Nestel D, Krogh K, Kolbe M. Exploring realism in healthcare simulations. Pp. 23–28 in Healthcare Simulation Education. Chichester, UK: John Wiley & Sons, Ltd, 2017.

66. Christian MS, Bradley JC, Wallace JC, Burke MJ. Workplace safety: A meta-analysis of the roles of person and situation factors. Journal of Applied Psychology, 2009; 94(5):1103–1127.

67. Edmondson AC, Lei Z. Psychological safety: The history, renaissance, and future of an interpersonal construct. Annual Review of Organizational Psychology and Organizational Behavior, 2014; 1:23–43.

68. van Dyne L, Ang S, Botero IC, Dyne L Van. Conceptualizing employee silence and employee voice as multidimensional constructs. Journal of Management Studies, 2003; 40(6):1359–1392.

69. O’Donovan R, McAuliffe E. A systematic review exploring the content and outcomes of interventions to improve psychological safety, speaking up and voice behaviour. BMC Health Services Research, 2020; 20(1):101.

70. King HB, Battles J, Baker DP, Alonso A, Salas E, Webster J, Toomey L, Salisbury M. TeamSTEPPS™: team strategies and tools to enhance performance and patient safety. P. in Henriksen K, Battles J, Keyes M, Grady M (eds). Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 3: Performance and Tools). Rockville, MD: Agency for Healthcare Research and Quality (US), 2008.

71. Maher RC. Application Example 2: Cockpit Voice Recorders. Pp. 137–142 in Maher RC (ed). Principles of Forensic Audio Analysis. Springer Nature Switzerland, 2018.

72. Fischer U, Orasanu J. Error-Challenging Strategies: Their Role in Preventing and Correcting Errors. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2000; 44(1):30–33.

73. Nevile M, Walker M. A context for error: using conversation analysis to represent and analyse recorded voice data. Human Factors and Aerospace Safety, 2005; (June):32.

74. Sassen C. Linguistic Dimensions of Crisis Talk. Amsterdam: John Benjamins Publishing Company, 2005.

75. Sexton JB, Helmreich RL. Analyzing Cockpit Communications: The Links Between Language, Performance, Error, and Workload. Journal of Human Performance in Extreme Environments, 2000; 5(1).

76. Orasanu J, Fischer U. Team cognition in the cockpit: Linguistic control of shared problem solving. Pp. 189–194 in Proceedings of the 14th Annual Conference of the Cognitive Science Society. New Jersey: Lawrence Erlbauw Associates, Inc., 1992.

77. Gurung G, Derrett S, Gauld R, Hill PC. Why service users do not complain or have ‘voice’: a mixed-

methods study from Nepal's rural primary health care system. BMC Health Services Research, 2017; 17(1):81.

78. Schwappach DLB, Gehring K. Trade-offs between voice and silence: a qualitative exploration of oncology staff's decisions to speak up about safety concerns. BMC health services research, 2014; 14(1):303.

79. Loewenstein GF, Weber EU, Hsee CK, Welch N. Risk as feelings. Psychological Bulletin, 2001; 127(2):267–286.

80. Leveson NG. Engineering a Safer World: Systems Thinking Applied to Safety. MIT press, 2011.

81. Reason J. Human error: models and management. British Medical Journal, 2000; 172(June):320–770.

82. Renn O. Concepts of risk: a classification. Pp. 53–79 in Krimsky S (ed). Social Theories of Risk. Praeger, 1992.

83. Viscusi WK, Zeckhauser RJ. Regulating Ambiguous Risks: The Less than Rational Regulation of Pharmaceuticals. Journal of Legal Studies (0047-2530), 2015; 44:S387–S422.

84. Slovic P. Perception of risk. Science, 1987; 236(4799):280–285.

85. Slovic P, Finucane ML, Peters E, MacGregor DG. Risk as Analysis and Risk as Feelings: Some Thoughts about Affect, Reason, Risk, and Rationality. Risk Analysis, 2004; 24(2):311–322.

86. Slovic P. Understanding perceived risk: 1978-2015. Environment, 2016; 58(1):25–29.

87. Laurent J, Chmiel N, Hansez I. Personality and safety citizenship: the role of safety motivation and safety knowledge. Heliyon, 2020; 6(1):e03201.

88. Murphy E, Dingwall R. Informed consent, anticipatory regulation and ethnographic practice. Social Science & Medicine, 2007; 65(11):2223–2234.

89. O'Connor P, Campbell J, Newon J, Melton J, Salas E, Wilson KA. Crew resource management training effectiveness: A meta-analysis and some critical needs. International Journal of Aviation Psychology, 2008; 18(4):353–368.

90. Helmreich RL, Merritt AC, Wilhelm JA. The evolution of crew resource management training in commercial aviation. The international journal of aviation psychology, 1999; 9(1):19–32.

91. O'Connor P, Jones DW, McCauley ME, Buttrey SE. An evaluation of the effectiveness of the crew resource management programme in naval aviation. International Journal of Human Factors and Ergonomics, 2012; 1(1):21.

92. Detert JR, Treviño LK. Speaking Up to Higher-Ups: How Supervisors and Skip-Level Leaders Influence

Employee Voice. Organization Science, 2010; 21(1):249–270.

93. Bienefeld N, Grote G. Speaking up in ad hoc multiteam systems: Individual-level effects of psychological safety, status, and leadership within and across teams. European Journal of Work and Organizational Psychology, 2014; 23(6):930–945.

94. Reason J. Human Error. Cambridge, UK: Cambridge university press, 1990.

95. Wiegmann DA, Shappell SA. A Human Error Approach to Aviation Accident Analysis. New York: Routledge, 2016.

96. Hofstede G. Dimensions Do Not Exist: A Reply to Brendan McSweeney. Human Relations, 2002; 55(11):1355–1361.

97. McSweeney B. Hofstede's model of national cultural differences and their consequences: A triumph of faith-a failure of analysis. Human Relations, 2002; 55(1):89–118.

98. Malloy D, Hadjistavropoulos, McCarthy E, Evans R, Zakus D, Park, Lee, Williams. Culture and organizational climate: nurses' insights into their relationship with physicians. Nursing ethics, 2009; 16(6):719–733.

99. Frosch DL, May SG, Rendle KAS, Tietbohl C, Elwyn G. Authoritarian Physicians And Patients' Fear Of Being Labeled "Difficult" Among Key Obstacles To Shared Decision Making. Health Affairs, 2012; 31(5):1030–1038.

100. Noland CM, Carmack HJ. Narrativizing Nursing Students' Experiences With Medical Errors During Clinicals. Qualitative Health Research, 2015; 25(10):1423–1434.

101. Schwartz MS, Dunfee TW, Kline MJ. Tone at the top: An ethics code for directors? Journal of Business Ethics, 2005; 58(1):79–100.

102. Detert JR, Burris ER, Harrison DA. Debunking four myths about employee silence. Harvard business review, 2010; 88(6):26.

103. Tailstrike. Cockpit Voice Recoder database. 2019.

104. Plane Crash Info. Last Words. 2019.

105. Aviation-Safety Network. Transcripts. 2019.

106. NATS. Analysis of Infringements in NATS-Controlled Airspace. 2013.

107. Erez M, Gati E. A dynamic, multi-level model of culture: From the micro level of the individual to the macro level of a global culture. Applied Psychology: an international review, 2004; 53(4):583–598.

108. Hofstede G. Dimension data matrix. 2015.

109. World Bank Group. Air transport, registered carrier departures worldwide. 2021.

110. Detert JR, Burris ER. Leadership Behavior and Employee Voice: Is the Door Really Open? Academy of Management Journal, 2007; 50(4):869–884.

111. Lin S-H, Johnson RE. A Suggestion to Improve a Day Keeps Your Depletion Away: Examining Promotive and Prohibitive Voice Behaviors Within a Regulatory Focus and Ego Depletion Framework. Journal of Applied Psychology, 2015; 100(5):1381–1397.

112. Barlow M, Watson B, Rudolph J. How being a great Receiver can change the game in speaking up conversations. MedEdPublish, 2019; 8(3):1–6.

113. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. Science, 1981; 211(4481):453–458.

114. Appelbaum SH. Socio-technical systems theory: an intervention strategy for organizational development. Management Decision, 1997; 35(6):452–463.

115. Guldenmund FW. The nature of safety culture: a review of theory and research. Safety Science, 2000; 34(1–3):215–257.

116. Mao LR. Beyond politeness theory: 'Face' revisited and renewed. Journal of Pragmatics, 1994; 21(5):451–486.

117. House RJ, Hanges PJ, Javidan M, Dorfman PW, Gupta V. Leadership, Culture, and Organizations: The GLOBE Study of 62 Societies., 2004.

118. Antonsen S. Safety culture and the issue of power. Safety Science, 2009; 47(2):183–191.

119. Weick KE. The Nature of Sensemaking. Sensemaking in Organizations, 1995:1–62.

120. Douglas M. Risk and Blame: Essays in Cultural Theory. Routledge, 1992.

121. Morrison EW, Wheeler-Smith SL, Kamdar D. Speaking up in groups: a cross-level study of group voice climate and voice. The Journal of applied psychology, 2011; 96(1):183–191.

122. Merritt A, Helmreich R. Human factors on the flight deck: the influences of national culture. Journal of Cross-Cultural Psychology, 1996; 27:5–24.

123. Blanco M, Clarke JR, Martindell D, Denise M. Wrong site surgery. AORN Journal, 2009; 90(7):814–816.

124. Johnson HL, Kimsey D. Patient Safety: Break the Silence. AORN Journal, 2012; 95(5):591–601.

125. Guevara G, Begault DR, Sunder K, Anderson M. Mitigation of Attentional Tunneling in the Flight Deck Using a Spatial Auditory Display. NASA Technical Reports Server (NTRS), 2018; (June).

126. Kitayama S, Park H, Sevincer AT, Karasawa M, Uskul AK. A cultural task analysis of implicit

independence: Comparing North America, Western Europe, and East Asia. Journal of Personality and Social Psychology, 2009; 97(2):236–255.

127. Shappell S a, Wiegmann D a. The Human Factors Analysis and Classification System – HFACS. USDOT/FAA/AM-00/7 Office of Aviation Medicine, 2000:19.

128. Aviation-Safety Network. Statistics by period. 2019.