Safety Listening in Organizations: An Integrated Conceptual Review

Alyssa M. Pandolfo (iD)
London School of Economics and Political Science, London, UK

Tom W. Reader
London School of Economics and Political Science, London, UK

Alex Gillespie
London School of Economics and Political Science, London, UK; Oslo New University College, Oslo, Norway

Abstract

Failures of listening to individuals raising concerns are often implicated in safety incidents. To better understand this and theorize the communicative processes by which safety voice averts harm, we undertook a conceptual review of "safety listening" in organizations: responses to any voice that calls for action to prevent harm. Synthesizing research from disparate fields, we found 36 terms/definitions describing safety listening which typically framed it in terms of listeners' motivations. These motivational accounts, we propose, are a by-product of the self-report methods used to study listening (e.g., surveys, interviews), which focus on listening perceptions rather than actual responses following speaking-up. In contrast, we define safety listening as a behavioral response to safety voice in organizational contexts to prevent harms. Influenced by cognitive, interactional, and environmental factors, safety listening may prevent incidents through enabling cooperative sensemaking processes for building shared awareness and understanding of risks and hazards.

Corresponding author:
Alyssa M. Pandolfo, Department of Psychological and Behavioural Science, Connaught House, 63-65 Aldwych, London, WC2B 4EJ, UK.
Email: a.pandolfo@lse.ac.uk

Plain Language Summary
Numerous global accidents (e.g., Tenerife air disaster), disasters (e.g., Challenger space-shuttle crash), and scandals (e.g., Enron's accounting scandal) stem from a shared cause: listeners failing to act upon legitimate voiced concerns. While research mainly centers on understanding and encouraging individuals to raise concerns, fewer studies explore listeners' responses. In this review, we advocate developing the concept of 'safety listening' (listeners' responses to voice aiming to prevent major harms). Our conceptual review positions safety listening as the necessary counterpart to safety voice; differentiates safety listening from other listening types (e.g., active listening); examines its terms/definitions, explanations, and measurements; and explores its influencers (e.g., listener motives) and impacts (e.g., injuries). Scholars have used 36 unique terms/definitions to describe safety listening (e.g., whistleblower retaliation) and often link it to listeners' motivations (i.e., they listened because they wanted to). This motivational focus, we argue, is a by-product of using self-report methods (e.g., surveys, interviews) that elicit perceptions rather than observing the actual behavioral responses to voice. To advance the literature, we propose using a standard definition, observing safety listening in real-life data (e.g., 9-1-1 calls, airline transcripts), and exploring mechanisms for how voice and listening cause outcomes (e.g., accidents). In sum, this review creates the foundation for future research to develop a comprehensive and cumulative understanding of safety listening and will ultimately contribute to preventing future accidents and scandals.

Theory on preventing institutional failures (e.g., accidents, scandals) has long emphasized the importance of stakeholders raising concerns to address and avert harm (Westrum, 2014). Accordingly, various research streams on speaking-up have emerged (Jones et al., 2021), which have not only primarily focused on safety voice (Noort et al., 2019), but also incorporated work on ethical voice (Chen & Treviño, 2023) and whistleblowing (Blenkinsopp et al., 2019). The unifying theme across such research is that individuals voice concerns to elicit action to address serious breakdowns in organizations and prevent harm (for parsimony, and due to a common focus on avoiding harm, we refer to such phenomena as "safety voice" from herein).

While safety voice acts are necessary to stop accidents and ensure safety, they are also insufficient: incident analyses often highlight how voice can go unheard before and during major institutional failures (Hald et al., 2020). Indeed, as illustrated in Table 1, ineffective listening is identified as a contributory factor in multiple failures, including the Tenerife air disaster (583 killed; Weick, 1990), the Bhopal disaster (500,000 exposed to toxic gas; Taylor, 2014), and the U.S. Gymnastics scandal (300 sexually assaulted; Kirby, 2018). Such cases illustrate that safety voice is only effective when followed by "safety listening."

Understanding why organizations' members fail to listen to legitimate and consequential concerns before avoidable failures is acknowledged as a critical knowledge gap in the organizational psychology, voice, whistleblowing, and healthcare literatures (Jones & Kelly, 2014; Vandekerckhove et al., 2014). We address this in our review, and, through its undertaking, develop a literature-based conceptualization of safety listening.

Background

Effective communication within organizations is crucial for ensuring that safety-related

Table 1. Example Organizational Disasters Preceded or Prolonged by Listening Failures

Disaster	Year(s)	Example voice attempt	Example listening failure	Consequences
Tenerife air disaster (Weick, 1990)	1977	The first officer of a KLM airplane observed that the captain had started takeoff without clearance and tried to alert air traffic control and other aircraft crews by saying a phrase that could be heard either as "We are at takeoff" or "We are, uh, taking off".	Recipients did not realize this was a warning and interpreted the ambiguous statement as "We are now at takeoff position".	The KLM collided into another airplane on the runway, killing 583 people.
Bhopal disaster (Taylor, 2014; Weick, 2010)	1984	One of the Union Carbide pesticide plant operators said "Hey can you smell it? I swear there's MIC [methyl isocyanate; a highly toxic gas] in the air".	The other operators replied, "It's not MIC you can smell, it's Flytox [mosquito spray]".	At least 30 tons of MIC were released, over 500,000 people were exposed, with an estimated death toll of 15,000 people.
Bundaberg Base Hospital scandal (Cleary & Duke, 2019)	2003–2005	The Nurse Unit Manager emailed the Director of Medical Services "as certain very disturbing scenarios have occurred" resulting from Jayant Patel (the Director of Surgery)'s behavior.	The Director of Medical Services did not respond to this email.	Patel was tried on three counts of manslaughter, pleaded guilty to concealing information which questioned his medical competence, and was barred from practicing medicine in Australia.
U.S. Gymnastics sex abuse scandal (*BBC News*, 2018; Dator, 2021; Mack & Lawler, 2017)	1994–2017	In 1999, an athlete told her coach and trainers that Larry Nassar (U.S. Gymnastics' physiotherapist) was penetrating her vagina during treatment for a hamstring injury.	The coach and trainers told her that Nassar was an Olympic doctor and knew what he was doing.	Nasser received a 60-year federal prison sentence for child pornography in 2017 and an additional 40–175 years on seven counts of criminal sexual assault in 2018.
Flint water crisis (CNN Editorial Research, 2021)	2014–2019	Flint residents brought jugs of discolored tap water to a community forum, complaining the water was contaminated and causing illness.	To assure residents that the water was safe, Flint Mayor Dayne Walling drank tap water on television.	Residents filed a federal class action lawsuit claiming that state and city officials knowingly exposed them to toxic water.

information is shared and acted upon (Westrum, 2014), with many models and theories focusing on linear message transmission from voicers to listeners. For instance, in their seminal communication model, Shannon and Weaver (1949) conceptualize dyadic communication in terms of voicers encoding and sending messages using specific channels (e.g., verbal), which may be negatively impacted by noise (e.g., static, disruptions), and is decoded by listeners. Likewise, at the organizational level, Westrum (2004)'s information flow theory posits that organizations with freely flowing risk-related information across levels are more effective in addressing problems than organizations characterized by withholding, distorting, or siloing information.

Scholars have argued that such communication conceptualizations primarily focus on message transmission (i.e., voice) rather than reception (i.e., listening; Macnamara, 2018; Vandekerckhove et al., 2014). Accordingly, as communication is an ongoing, interactive, and iterative process where individuals interpret and develop the topic being communicated, scholars (e.g., Schramm, 1955) have extended the Shannon and Weaver model to include feedback from receivers to senders. For example, one feedback type may be "third turn repairs" where listeners try to confirm understandings and/or correct misunderstandings between interlocutors (Schegloff, 1992).

Like the wider communication literature, research on safety in teams and organizations has primarily explored how safety-related information (e.g., about hazards) is shared, rather than how it is listened to. Notably, safety research has focused on safety voice, which is generally conceptualized in terms of discretionary, typically upward communication where individuals raise significant concerns about potential harms to those who can intervene or escalate the problem. Safety voice studies have typically examined speaking-up with safety-related information (Noort et al., 2019), or incorporated related concepts like employee voice (Morrison, 2014); whistleblowing

(Near & Miceli, 1985); voice in action teams (Krenz & Burtscher, 2021); and ethical voice (Chen & Treviño, 2022).

As safety voice is considered essential for ensuring information flow about hazards, risks, and potential safety improvements, investigations have generally focused on speaking-up's antecedents (Chamberlin et al., 2017). Common findings are that individuals often engage in voice acts depending on their attitudes and skills for managing safety (Salas et al., 2020) and psychological safety where they believe speaking-up will lead to change (Morrison, 2014) and will not result in negative consequences (Edmondson, 2018). However, while safety voice is an increasingly well understood phenomenon (Bazzoli & Curcuruto, 2021), the mechanism by which it prevents accidents and harms—namely, safety listening—is less examined (Vandekerckhove et al., 2014). This is a significant gap: it is essential to understand the factors that determine how and whether voice is heard, understood, and responded to (Barlow et al., 2019) to explain how safety voice can best achieve its goal of eliciting action. While research outside of safety has provided insight on how listening in organizations can be encouraged (e.g., by signaling aids; Itzchakov & Kluger, 2017), little research has focused on dynamic and high-stakes contexts (e.g., aviation emergencies), or how safety voice invokes actions for preventing accidents and avoiding harm (e.g., raising alarms, instituting changes).

In summary, while there has been extensive research on improving safety voice in organizations, there is no established safety listening counterpart. Although the workplace listening literature has focused on responses to voice acts that might be considered routine (e.g., listening to performance evaluation concerns), the nature of listening to voice where serious failures might be averted has not been studied. Indeed, the broader psychology literature does not indicate precisely how listening to safety voice should be conceptualized: for instance, researchers have diversely conceived listening as being how receivers perceive voicers,

observable (e.g., nodding) and unobservable (e.g., comprehension) behaviors in response to voice, feelings about voice (e.g., exhaustion), intentions based on listening to voice, and voicers' perception of being heard (Kluger & Izchakov, 2022; Yip & Fisher, 2022).

We conducted a conceptual review of research investigating safety listening in organizations. Our goal was, through synthesizing research and observations on how safety voice is listened to and leads to outcomes, to conceptualize the nature, antecedents, indicators, and outcomes of safety listening. For example, in terms of describing what safety listening is (e.g., behaviorally), understanding what drives it (e.g., attitudes, culture), exploring how it should be measured (e.g., naturalistic data), and considering its relationship with outcomes (e.g., how it prevents accidents). Guided by the idea that safety voice is a call to action and a successful listening episode must result in some action (e.g., stopping take-off; Noort et al., 2021a), our initial conceptualization of safety listening was: listeners' behavioral responses to safety voice acts to address potential harms in organizational contexts.

Our review's contribution is to develop foundational principles for the concept of safety listening, outline how it can be advanced theoretically and empirically, and identify how it can improve psychological research on topics like safety and ethics in organizations.

After outlining our search methods, we describe how researchers have conceptualized and investigated safety listening, including its distinctions from other listening types, its terms/definitions, its theoretical explanations, its measurements, and its pragmatics/antecedents/outcomes. To integrate research on safety listening into the wider psychology literature, we interpreted our findings using the Shannon and Weaver model, third turn repairs, and Westrum's information flow theory. Through critically evaluating the literature—for instance, in relation to fragmented terms/definitions, a focus on listeners' motivations, and the overuse of self-report measures—we recommend avenues for future investigation.

Methods

We undertook a robust literature search to identify safety listening conceptualizations. Our conceptual review's purpose was to synthesize and reconcile fragmented terminologies, definitions, and measurements of safety listening to aid conceptual demarcation and theory development (Hulland, 2020). Accordingly, rather than performing an exhaustive literature search, we aimed to understand how scholars have conceptualized safety listening.

First, we performed a systematic literature review following Siddaway et al. (2019)'s recommendations. Using Scopus, Web of Science, and PsycINFO, we searched for publications, which included listening, concerns, failure, and organizations (or synonyms) within abstracts (Web of Science and PsycINFO) and titles, abstracts, or keywords (Scopus; Table 2 for the search strategy). We verified our search strategy by ensuring that it included five articles central to our research (i.e., Hald et al., 2020; Harlos, 2001; Jones & Kelly, 2014; Martin et al., 2021; Peirce et al., 1998). As listening is implicit to many literatures (e.g., teamwork, psychological safety), we included publications that explicitly theorized or investigated responses to high consequence concerns (Table 3 for inclusion criteria). Alongside focusing on safety voice studies, we considered other potentially relevant research: for example, investigations on voice more generally, whistleblowing, speaking-up about problems in healthcare or dangerous workplace behavior (e.g., sexual harassment), and unethical conduct. AMP (research psychologist) conducted searches, removed duplicates, and screened results based on titles, abstracts, and full texts in March 2022 (Figure 1 for the flowchart diagram). We conducted an interrater reliability assessment where a psychology PhD candidate screened 30 randomly selected publications, resulting in a Cohen's kappa of 0.86 (substantial strength of agreement). All authors regularly and collectively discussed borderline case inclusion and interpreted findings.

Table 2. Systematic Review Search Strategy

Electronic search	Scopus, Web of Science, PsycINFO
Keywords	Title, Abstract, or Keywords (Scopus) or Abstract (Web of Science, PsycINFO) contain: (respon* OR listen*) AND (concern* OR complain* OR whistleblow* OR whistle-blow*) AND (disaster* OR accident* OR fail* OR crisis* OR safety) AND (organi*ation* OR team*)
Limitations[a]	Articles, book chapters, review articles[b] English Subject areas of business, psychology, or social sciences[c]

[a]For Scopus, we manually applied these at the title screening stage for remaining databases.
[b]We excluded unpublished work and did not apply date parameters to our search, as we were interested in the literature's existing conceptualizations.
[c]As we were interested in how organizational psychology conceptualizes responses to high-consequence concerns in organizations, we initially focused on social science, business, and psychology publications and broadened this criterion in the hand search to include healthcare articles.

Next, we conducted two manual searches in March–June 2022 and June–July 2023 because our initial search strategy excluded relevant papers due to terminology variations (e.g., "cockpit crew" instead of "team"). First, we examined publications citing and cited by those included in our systematic review and hand search. Second, we searched journals likely to contain relevant publications (e.g., *International Journal of Listening*). Finally, we investigated key authors' Google Scholar pages and reviews in the following domains for safety listening research: safety voice (Noort et al., 2019), employee voice (Lainidi et al., 2023; Morrison, 2011, 2014), voice in action teams (Krenz & Burtscher, 2021), whistleblowing (Blenkinsopp et al., 2019; Mesmer-Magnus & Viswesvaran, 2005), ethical voice (Chen & Treviño, 2023), workplace listening (Kluger & Izchakov, 2022; Yip & Fisher, 2022), safety culture (Bisbey

et al., 2021), teamwork (Salas et al., 2020), and ethical leadership (Brown & Treviño, 2006). We recognize that publications post-July 2023 were not included, and others may have been missed despite our multiple search strategies.

Our review is contextualized against recent workplace listening reviews (Kluger et al., 2023; Kluger & Izchakov, 2022; Yip & Fisher, 2022)—it is narrower in topic (i.e., safety listening) yet broader in scope because these reviews focused on the management, communication studies, and psychology literatures. AMP and a psychology MSc postgraduate research assistant assessed listening-related terminologies and definitions; explanations (i.e., theories, models, processes); methods and data sources; and pragmatics, antecedents, and outcomes for each publication.

Findings

We identified 57 articles, published between 1982 and 2023, focusing on listening to high consequence concerns in organizational contexts. Of these, 43 were empirical (comprising 46 studies), 11 were theoretical, and three were reviews. Table 4 synthesizes our findings into a set of key observations, critiques, and recommendations.

Safety listening conceptualizations

Safety listening's distinguishability. In developing the concept of safety listening, our first goal was to distinguish it from other listening forms (e.g., active listening). It is distinguishable in four ways. First, it can result in significant and sometimes life-or-death outcomes (Hällgren et al., 2018), including physical (e.g., fatalities), psychological (e.g., trauma), and environmental (e.g., pollution) harms.

Second, safety listening has been studied in risky (e.g., aviation; Noort et al., 2021a) and emergency (e.g., healthcare; Long et al., 2020) contexts, though it may also occur in disrupted contexts (e.g., bomb in subway stations; Hällgren et al., 2018). Such contexts include

Table 3. Inclusion Criteria

Domain	Include	Exclude	Rationale

Publication	Peer-reviewed articles
	    Literature reviews
	    Book chapters
	    Case studies
	    Commentaries
		Duplicates
		    Book/film reviews
		    Dissertations
		    Periodicals
		    Editorials
		    Ongoing/unpublished research
		    Errata
		    Conference proceedings
			We included peer-reviewed, published research

Language	English	All other languages	We did not have the resources to translate non-English articles

Topic	Listeners' behavior responding to safety voice demanding action to prevent harms in organizational contexts	Speaking up (not responses)
		    Antecedents of speaking up only (e.g., soliciting voice; voice climate)
		    Perceptions of voicers (e.g., as dissidents)
		    Speaking up outcomes to the group or organization (e.g., turnover, performance)
		    Responding to wrongdoing or hazards directly (i.e., no voice preceding response)
		    Recommendations for how to listen
		    Non-organizational complaints (e.g., between couples in marital therapy)
		    Incident reporting systems
		    Auditory problems or diseases
			We included publications which conceptualized responses to safety voice in organizations

Concern type	Safety and/or environmental hazard(s) (e.g., poor patient care) Security or terrorism threats Institutional prejudice (e.g., homophobia, racism, brutality) Bullying, harassment, abuse, rape Fraud, theft Workplace misconduct	Suggestions of improving organizational strategy (e.g., profitability) Organizational citizenship behaviors Low-consequence or mundane complaints (e.g., disliking paint color)	We included publications which had consequential written or spoken messages—i.e., those which may result in safety-related outcomes, including physical, mental, emotional, social, and/or environmental harms
Voicer	characteristics	Organization employee, team, or group (e.g., union)	N/A

Table 3. (continued)

| Domain	Include	Exclude	Rationale
		Organization customer/client (e.g., patients, bus passengers) External organization	
We intentionally included a range of voicers to cast a wide net			
Listener	characteristics	Organization (e.g., employee, team, group) Individuals acting on behalf of an organization (e.g., spokesperson, consultant) Intended recipient of voice	Non-employees Third parties (e.g., co-worker overhearing voice)
We limited listeners to those within	organizations because they would be likelier to address complaints		
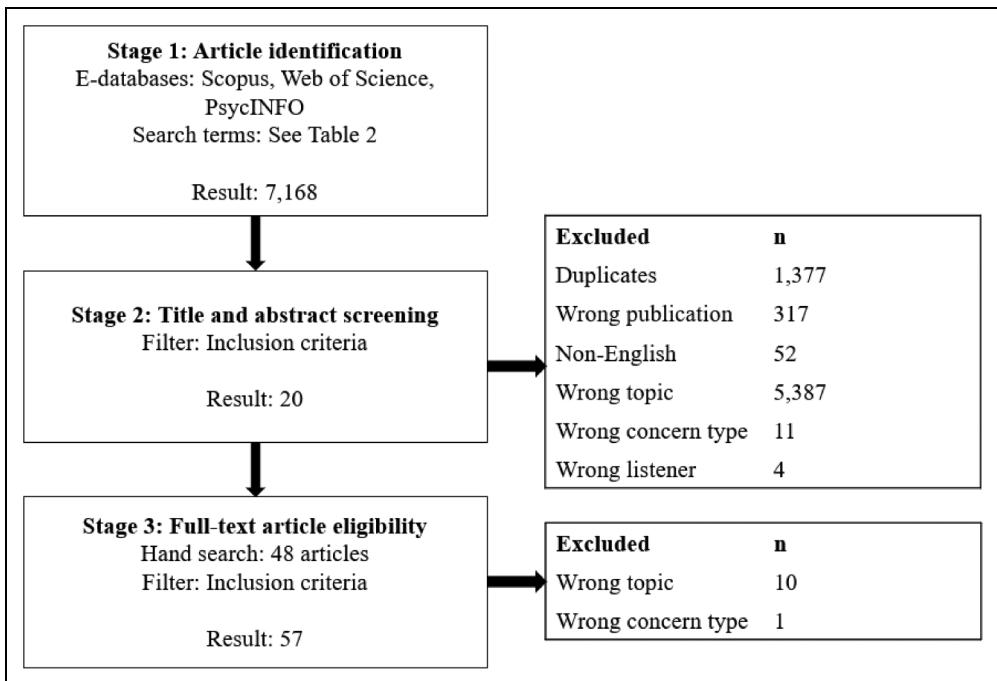
Figure 1. Flow diagram of the systematic literature search on safety listening.

Table 4. Key Conceptual Review Findings

Finding	Description	Critique	Recommendations for future research
1. The concept of safety listening is fragmented within and between literatures.	We found 36 unique terms/definitions of safety listening.	The extent to which these terms overlap or are distinguishable is unclear.	Use a standardized concept and definition, namely "safety listening," defined as listeners' behavior responding to safety voice demanding action to prevent harms in organizational contexts.
2. Scholars have predominantly explained safety listening in terms of listeners' attitudes and motivations.	Poor safety listening has been viewed as strategic, deliberate, and because "listeners did not want to listen".	While motivations influence listening behaviors, concentrating on listeners' attitudes narrows our understandings.	Develop a conceptual model and theory which explain the organizational and system dynamics through which safety voice and listening impact outcomes (e.g., accidents).
3. Scholars have mainly assessed safety listening using proxies of behavior.	Studies have commonly measured safety listening using self-report methodologies (e.g., surveys, interviews), generally assessing the voicer's or listener's perspective of the conversation.	Self-report measures are susceptible to biases, occur in decontextualized conditions, and over-rely on participants' imaginations.	Assess safety listening behaviors using hitherto neglected naturalistic data (see Tables 7 and 8 for more detail). This will enable a more balanced approach to understanding safety listening.
4. Scholars conceptualize effective safety listening as agreeing with voicers and making them feel heard.	Researchers assume that "better" safety listening is endorsing the voice message, making voicers feel heard, and addressing problems. Conversely, "ineffective" listening is ignoring or disagreeing with voicers.	It is likely that effective safety listening is more nuanced than agreement. For instance, disagreement may be effective listening if voicers are wrong. Additionally, making voicers feel heard may be ineffective safety listening if the listener fails to address the problem at hand.	Establish a taxonomy of safety listening behaviors, which can then be used to create a reliable and valid coding framework. Assess safety listening behaviors rather than self-reports.
5. Scholars generally position the voice/listening/outcome relationship as one-shot.	Publications typically conceptualized safety voice as occurring, then one act of safety listening, then	The voice/listening/outcome relationship is likely iterative. Safety voice begins a conversation, safety	Include a feedback loop between safety voice and safety listening in theoretical models. Recognize that effective

(continued)

Table 4. (continued)

Finding	Description	Critique	Recommendations for future research
	outcomes (e.g., scandal).	listening can further the conversation by prompting sense-making or inviting future voice acts. Sometimes listeners are unable to address complaints and may become future voicers.	safety listening may include sense-making.
6. Scholars typically investigate clear-cut situations.	Studies have often examined situation where voicers have a clear complaint and know who to voice to.	Voicers may have fuzzy concerns and organizations may have multiple and unclear pathways for raising concerns (e.g., informal and formal complaint channels).	Understand how safety listening is affected when voicers wish to raise concerns but are unsure how or when voicers raise concerns to improper recipients.
7. Many studies do not specify complaints' nature.	Publications did not distinguish responses to complaints, which were verbal vs. written. They also did not often specify whether technology was used to deliver complaints (e.g., email).	Non-verbal information from face-to-face or verbal communication may influence understandings. Having a written record of concerns may prompt a response. Certain technologies may prompt instant (e.g., phone call) or delayed safety listening (e.g., email).	Specify the nature of complaints and explicitly investigate whether complaints in different channels are responded to differently.

formalized organizational settings with distinct role definitions and imbalances in roles, responsibilities, and expertise between voicers and listeners (e.g., nurses voicing to doctors; McDonald & Ahern, 2000), and dynamic environments where events cannot be fully rehearsed (e.g., security).

Third, listeners often make decisions amidst high cognitive loads, stress, urgency, and danger (e.g., military). Listeners face a dilemma: acting on voice (e.g., cancelling take-off following technical concerns) could lead to proximal consequences (e.g., disruption) yet potentially prevent future incidents while inaction may create distal risk (e.g., accidents). Likewise, erroneous actions may be challenging to reverse (e.g., shutting down the wrong engine; Krenz & Burtscher, 2021). Thus, listeners must engage with concerns, compare risks tied to different actions, and determine appropriate actions in uncertain conditions.

Last, safety listening is verifiable. As safety voice requests novel or corrective action by listeners (e.g., stopping harassment; Peirce et al., 1998), safety listening is observable in listeners' responses (e.g., creating plans; Groves et al., 2021), team-members' shared understandings of the situation, and situations' outcomes (e.g., de-activated bomb following citizens' concerns). Table 5 outlines safety listening's defining features, including its content, observability, and context.

Table 5. Safety Listening Definition and Defining Features

Safety listening: Listeners' behavior responding to safety voice demanding action to prevent harms in organizational contexts.

Defining feature	Description
Safety-related message content	What is listened to pertains to current or future hazards, misconduct, or wrongdoing, which may result in life-or-death outcomes and significant physical, social, or environmental harms.
Requires and follows safety voice	Safety listening is the counterpart to safety voice (i.e., raised concerns which could result in significant harms), where safety voice is necessary but insufficient to address concerns.
Listeners are humans	Safety listening is done by humans after receiving safety voice messages interpersonally or via organizational listening architectures (e.g., formal complaint reporting channel).
Observable in listeners' behavior	Safety listening is verifiable in listeners' communications and/or actions after safety voice. For instance, listeners can address or dismiss the concern, begin sense-making, or ask for clarification. Such behavior may then reinforce, extinguish, or punish future instances of safety voice.
Often iterative	While safety listening can be one-shot, it often prompts future voice and creates trajectories of voice/listening acts.
Can occur at the dyad, team, or organizational level	Safety listening can occur between two or more team members (e.g., surgery team before an operation) but also at the organizational level (e.g., complaints department).
Embedded in organizational and social contexts	Safety listening occurs in risky, emergency, and disrupted organizational contexts and it is influenced by listeners' cognitions and skills, interpersonal factors, and structural factors.

Safety listening's fragmentation. To test the idea that safety listening is a concept that is implicitly recognized as important within the literature but has not been crystalized into a defined term or phenomenon, we examined how listening was defined across the included studies. Publications used 36 unique terms/definitions for listening, with many seeming overlapping or with unclear delineations. For example, scholars used "retaliation" (Rehg et al., 2008), "whistleblower retaliation" (Kenny et al., 2019), and "official" and "unofficial reprisals" (McDonald & Ahern, 2000) to describe negative consequences following voicing. Similarly, Jones and Kelly (2014) renamed the "deaf effect" (Cuellar et al., 2006) as "organizational disregard" while retaining the same definition.

While some terminologies conveyed neutral or positive connotations (e.g., "reaction to speaking up"; Lemke et al., 2021), most illustrated consequences following voice (e.g., "retaliation"; Rehg et al., 2008) or exclusively negative responses (e.g., "silencing"; Tiitinen, 2020). Moreover, some terms were framed at the dyadic level (e.g., "receiver response"; Long et al., 2020), while others pertained to organizational dynamics (e.g., "organizational silencing"; Fernando & Prasad, 2019).

Safety listening's motivational framing. We next investigated how the literature has explained safety listening, finding that terminologies and definitions generally frame it as motivational. Terms such as "willful blindness" (Cleary & Duke, 2019) and "deaf ear syndrome" (Peirce et al., 1998) insinuate that inaction is intentional; listeners choose to turn blind eyes or deaf ears following concerns. Likewise, publications' definitions primarily focused on listeners' (in)actions following voice, including

addressing concerns (Tucker & Turner, 2015), ignoring complaints (Cleary & Duke, 2019), and retaliating (Rehg et al., 2008). Most definitions implied actions were deliberate (e.g., listeners' "willingness to refrain from […] retaliation"; Vandekerckhove et al., 2014, p.300) to pursue objectives (e.g., preventing further whistleblowing; Tiitinen, 2020). Such intentionality seems consistent with the behavioral literature (Skinner, 1963) where listeners' actions (e.g., punishment) may serve to extinguish future voice behaviors.

Safety listening explanations generally posit that listeners intentionally choose responses based on their attitudes, strategy, or motives. Following voice, listeners are thought to determine responses after rationally assessing whether problems exist, whether they are responsible for them, and whether they can address them (Pierce et al., 2004). Applying the theory of planned behavior, Vandekerckhove et al. (2014) posit that listeners are likelier to retaliate if they possess negative whistleblowing attitudes, improvable subjective norms (e.g., witnessing others retaliate), and poor perceived behavioral control (e.g., believing they cannot help). Similarly, Near and Miceli (1985) argue that, following whistleblowing, organizations determine whether the misconduct should cease and/or the whistleblower should be punished; choosing retaliation is seen as proportionate to the organization's dependence on the wrongdoing and inversely related to the whistle-blower's power (Miceli et al., 2008).

Research often frames poor listening as intentional. Ineffective responses are viewed as intended to silence or discredit voicers (Fernando & Prasad, 2019), highlight voicers' out-group membership (Barlow, 2021), rid listeners' negative emotions (Sumanth et al., 2011) and cognitive dissonance (Atkinson et al., 2022), prompt voicer conformity (Wellman et al., 2016), and minimize reputation loss (Near & Jensen, 1983). Moreover, Martin and Rifkin (2004) and Roulet and Pichler (2020) conceptualize whistleblowing responses as strategic maneuvers within games ("organizational jiu-jitsu" and "blame games",

respectively), designed to minimize personal and organizational culpability. However, explaining listening as motivational and strategic may overlook alternative explanations, including misunderstandings (Schegloff, 1992).

Safety listening methodologies

Having established how safety listening tends to be conceptualized and understood in organizational contexts, we investigated how it is studied, with, given voice's aim of soliciting action, a particular interest on whether and how it is studied behaviorally. Like safety voice (Noort et al., 2019), the 46 empirical safety listening studies occurred most frequently in healthcare contexts (n = 14) and in America (n = 19).

Considering that safety listening, ultimately, is about responding to voice acts aiming to elicit action, only two directly measured safety listening behaviors in naturalistic contexts. Lemke et al. (2021) observed voice and listening in teams administering pre-surgery anesthesia. They created behavioral codes assessing listeners' verbal, behavioral, and affective responses, and coded behaviors *in situ*. Noort et al. (2021a) analyzed behavioral trace data from conversations preceding airplane crashes between 1962 and 2018. They identified safety listening in the three conversational turns following safety voice, classifying it as immediate action, verbal affirmation, ignoring, or disaffirmation. The remaining studies used self-reported measures (e.g., surveys) or measured behavior in contrived settings (e.g., experiments); Table 6 details these methodologies. Notably, almost half of the studies (20/46) did not specify whether complaints were spoken or written, and few explicitly investigated technology-mediated complaints (e.g., phone calls, emails).

In sum, like voice (Lainidi et al., 2023) and other listening types (Kluger et al., 2023), safety listening insights were primarily obtained through measurements assessing self-reported imagined or recalled responses to high consequence concerns. Despite variously

Table 6. Safety Listening Assessments

Methodology	Perspective	*n*	Example
Survey	Voicer (voicing's consequences)	14	Parmerlee et al. (1982) asked whistleblowers to indicate whether they had been threatened with or had experienced specific forms of retaliation (e.g., exclusion from meetings previously attended; poor performance appraisal; demotion).
	Voicer (whether they believed they were heard)	1	Tucker and Turner (2015) asked young workers to indicate their agreement with items like "My supervisor cares about my safety opinions".
	Listener	2	Kundro and Rothbard (2023) asked online survey platform participants to indicate whether they retaliated against an employee who morally objected, using items like "I retaliated against this individual".
Interview	Voicer	5	Fernando and Prasad (2019) asked academics to describe what happened after complaining about sexual harassment at work.
	Listener	2	Groves et al. (2021) interviewed nurses about how they respond to patient safety concerns.
	Third parties	1	Peirce et al. (1998) interviewed attorneys who had litigated several hundred sexual harassment cases, asking them to discuss how firms "turned a deaf ear" toward sexual harassment complaints.
	Multiple perspectives	5	Kenny et al. (2019) interviewed whistleblowers and psychiatrists who worked with whistleblowers about whistleblowers' mental health following experiencing retaliation.
Vignette	Listener	8	Pierce et al. (2000) asked sheriff department employees to read a vignette depicting an employee who lodged a sexual harassment complaint against another employee who persistently rubbed her neck and made sexual jokes. Participants imagined they were human resources managers and rated the appropriateness of actions following the allegation (e.g., suspension, offering company-funded counseling). It is unclear whether sheriff department employees responding as human resource managers to hypothetical vignettes would accurately reflect human resource managers' actual responses.
Online experiment	N/A	5	Wellman et al. (2016) tasked university students to evaluate a business case with virtual "team-members" (in reality, scripted messages delivered by confederates). One "team-member" refused to finish their portion, citing ethical concerns. Participants rated their intent to engage in various behaviors toward this team-member (e.g., bullying) and wrote feedback messages on

(continued)

Table 6. (continued)

Methodology	Perspective	*n*	Example
			their performance. Independent coders rated these messages' content.
Case study	N/A	4	Reader (2022) analyzed a public inquiry report containing narratives from patients/relatives and staff members who experienced or delivered care at the Mid Staffordshire NHS Foundation Trust following its patient neglect scandal. Reader classified staff responses to concerns as accepting, denying, ignoring, or not acting.
Simulation	N/A	1	Barlow et al. (2023b) evaluated safety listening through medical simulations involving clinicians enacting a discharge scenario. A confederate voiced the scripted concern that a patient would be discharged without suitable home care. Researchers then assessed whether listeners acknowledged concerns, reassured voicers, and modified their language to facilitate understanding. Notably, participants' awareness that the simulation was related to their speaking-up training program may have influenced responses.
Naturalistic observations	N/A	1	See Lemke et al. (2021) example in-text.
Behavioral trace data	N/A	1	See Noort et al. (2021a) example in-text.

operationalizing safety listening, researchers have generally positioned "better" listening as agreeing with voicers, making voicers feel heard, and addressing problems (Barlow et al., 2023b; Noort et al., 2021a; Reader, 2022). Studies typically framed voice and listening as one-shot, studying instances where voicers knew how to effectively raise concerns and listeners could address them. Scenarios where voicers were uncertain how to complain and where listeners could not act (e.g., about a third party's error) were rarely examined.

Safety listening findings

To establish the existing corpus of safety listening knowledge and identify potential areas for future investigation, we explored observations on the following: i) safety listening's pragmatics and how it manifests in organizations, ii) safety listening's antecedents (e.g., preventative and promotive factors), and iii) safety listening's outcomes. When evaluating publications' findings, we considered communication repairs and organizational information flow.

Safety listening pragmatics. Three studies explored how listeners responded to voice, typically conceptualizing effective listening as agreement. Self-reported (i.e., interviews) and observed (i.e., simulated behavior) "appropriate" listening included acknowledgment, thanking voicers for speaking-up, and validating emotions (Barlow et al., 2023b; Groves et al., 2021). Conversely, responses using task-based questioning (e.g., "What actions are necessary to discharge this patient?") hindered listening because listeners did not address voicers' concerns. In contrast, Lemke et al. (2021) observed that responses were often neutral or validating, primarily comprised of short approvals or detailed explanations. While Groves et al.

(2021) conceptualized voice and listening as one-shot (i.e., voice, then listening, then outcome), Barlow et al. (2023b) and Lemke et al. (2021) (assessing naturalistic or simulated behavior) described listening as iterative; it furthered the conversation by prompting sense-making or inviting future voice acts. Notably, all studies were in healthcare contexts; it may be that listening pragmatics differ in other situations.

Safety listening antecedents. Publications measured antecedents at three levels: listeners' cognitive/skill-based factors, interactional dynamics among voicers and listeners, and structural factors within organizations. These studies quantitatively investigated listening's antecedents and outcomes or have qualitatively described antecedents and outcomes using institutional failure examinations and voicer/listener interviews.

Cognitive/skill-based factors

Publications explored listeners' motivations as a listening antecedent. Ineffective listening was preceded by listeners' poor motivated reasoning (Cleary & Duke, 2019), negative attitudes toward complaints (Hsieh et al., 2005), and expectations that voicers would address problems themselves (Wilkinson et al., 2015). Additionally, inadequate listening (e.g., retaliation) was likelier when listeners perceived voicers as threatening (Kenny, 2019), cold, or unlikeable (Wellman et al., 2016). Receivers also were unlikely to listen if they were stressed (Long et al., 2020) or feared disciplinary actions (Martin et al., 2021).

Complaints' perceived legitimacy also influences safety listening. Effective listening is encouraged when listeners receive compelling evidence (Mesmer-Magnus & Viswesvaran, 2005) and believe voicers were not responsible for incidents (Pierce et al., 2004). Yet, listening may be hindered if listeners excessively confirm complaints' legitimacy instead of focusing on comprehending and resolving issues (van Dael et al., 2022).

Inadequate listening skills may underpin ineffective listening. Some listeners indicated not knowing how to respond to voice (Barlow et al., 2023a) and having insufficient training to do so (Hsieh et al., 2005). Conversely, trained managers with experience with whistleblowers were likelier to safeguard them from retaliation (Vandekerckhove et al., 2014). A listening skill which would benefit from more investigation would be identifying and correcting misunderstandings (e.g., the voicer incorrectly perceived a problem when one did not actually exist)—but this would require looking beyond one-shot voice acts.

Interactional factors

Interactional dynamics may shape listeners' cognitions and responses, with research focusing on voicer/listener power and team members' support. Listening is likelier when there are low voicer/listener power disparities (Miceli et al., 2008)—for example, if voicers had high-status positions (Cortina & Magley, 2003), their roles included whistleblowing (e.g., auditing; Casal & Zalkind, 1995), and listeners respected voicers' seniority (Long et al., 2020). Conversely, ineffective listening was often preceded by hierarchical and expertise differences (e.g., junior pilots voicing to captains; Noort et al., 2021a). In short, listeners may have preconceptions as to who might have relevant safety information, which reduces their openness to voices from unexpected sources.

Third-party support may discourage retaliation. Voicers were unlikely to experience retaliation if colleagues understood why they voiced (Park et al., 2020), and voicers had supervisor (Mesmer-Magnus & Viswesvaran, 2005) and/or top management support (Miceli et al., 1999).

Structural factors

Interactional factors are influenced by structural factors, including reporting channels, organizational cultures, and organizational characteristics. Scholars distinguish between

different reporting channels: external (e.g., to regulators), informal internal (e.g., open-door policies), and formal internal (e.g., complaints systems). All are associated with ineffective listening. Using external whistleblowing channels increased retaliation's likelihood and severity (Mesmer-Magnus & Viswesvaran, 2005). Likewise, voicers using informal internal reporting channels believed these inadequately addressed complaints (Harlos, 2001). Ineffective formal reporting channels were poorly demarcated, bureaucratic, inadequately captured complaints' nuances, and prioritized achieving performance targets (e.g., reducing complaint numbers) over addressing complaints (Martin et al., 2021; van Dael et al., 2022). Studies have typically investigated "clear-cut" speaking-up where voicers have straightforward concerns, know how to navigate multiple channels, and have a single intended audience.

Organizations with poor safety cultures (Reader, 2022), cultures prioritizing performance over safety and ethics (Wilkinson et al., 2015), and cultures where complaints conflict with taken-for-granted assumptions (Hald et al., 2020) are associated with ineffective listening. Additionally, complaints were likelier to be disregarded in small firms, family-owned enterprises, decentralized or multinational setups, male-dominated sectors, and rural locations (Peirce et al., 1998). Organizations with ineffective listening often exhibited limited information sharing (Hsieh et al., 2005), possessed poorly defined policies concerning wrongdoings (Peirce et al., 1998), and excluded employees from problem-solving processes (Tiitinen, 2020).

In sum, consistent with Westrum (2014), structural factors including poor reporting channels and pathological or bureaucratic organizational cultures blocked information flow within organizations.

Safety listening outcomes

Five studies investigated consequential outcomes following safety listening using surveys, interviews, and behavioral trace data. Ineffective listening was associated with increased injuries (Tucker & Turner, 2015), airplane damage (Noort et al., 2021a), and death (Hald et al., 2020). Experiencing retaliation also worsened voicers' physical (Cortina & Magley, 2003) and mental health (Kenny et al., 2019).

Voicers who experienced retaliation also reported diminished job satisfaction (Cortina & Magley, 2003), supervisory relationships (Rehg et al., 2008), and career advancements (McDonald & Ahern, 2000). Voicers receiving multiple retaliation forms were likelier to report retaliation (Near & Miceli, 1986), and women who experienced retaliation were likelier to whistle-blow externally (Rehg et al., 2008). Certain listening behaviors extinguished future voice behaviors: voicers were reluctant to voice again when listeners provided extensive explanations (Lemke et al., 2021) and when senior listeners disaffirmed junior voicers (Noort et al., 2021a).

Discussion

Our review found research on safety listening to be highly fragmented in terms of definitions and conceptualizations and to be overly focused on attitudes toward listening. We concluded that, because safety listening influences outcomes through responding to voice acts, its conceptual basis lies in consequential (and thus observable) behavior (i.e., action) and its outcomes (e.g., harm avoidance), rather than listeners' or voicers' attitudes. From this standpoint, safety listening can be understood as a "world-making" behavior (Power et al., 2023), because what listeners say and do after hearing safety voice—for example, addressing concerns, correcting misunderstandings, and starting sense-making processes—determines how and whether action is taken to prevent accidents and avoid harm.

Like the communication literature, for example, Shannon and Weaver's one-shot communication model (i.e., voice-listening-outcome), safety listening research tended to consider only "one turn" of communication, and not how individuals iterate and sense-make across many turns to

understand safety concerns. For example, through third turn repairs, which are important for correcting misunderstandings in how listeners have understood voice acts (Noort et al., 2021a). Accordingly, like extensions to the Shannon and Weaver model (Schramm, 1955), we suppose that safety listening should not be considered a ballistic one-shot behavior, but rather part of an iterative sense-making process, with feedback loops between voice and listening.

Our review has implications for Westrum (2004)'s information flow theory. Like Westrum, we found that structural factors (e.g., improvable organizational cultures) block organizational information flow, sometimes with deleterious outcomes. Yet, Westrum focused on whether information was being voiced within organizations and when considering listeners' roles, primarily explored listeners' motivations upon hearing problems (e.g., preoccupation with power). As such, information flow theory can be broadened to incorporate misunderstandings and miscommunications to offer a more comprehensive perspective, including safety in uncertain conditions. Moreover, empirical studies should apply information flow theory at the moment of voice and listening to enhance this theory's applicability.

For the safety literature, our safety listening conceptualization is significant because it explains how different forms of voice—for instance that communicating promotive (e.g., safety-related improvements) or prohibitive voice messages (e.g., safety complaints)—can change the status quo, and that there are factors which underlie this (e.g., individual, contextual). For prohibitive voice, listening is essential for individuals and teams to form a shared situation awareness with the voicer, and thus act (or otherwise; Endsley, 1995). For promotive voice, how listening shapes outcomes is less clear, is likely to occur in a less urgent context, and may require more ongoing and iterative processes (e.g., on deciding whether to improve a safety procedure). Following the behavioral literature, for example, on operant conditioning (Skinner, 1963) and melioration theory (Herrnstein &

Prelec, 1991), it would also be valuable to understand how safety listening influences safety voice (e.g., through encouraging or extinguishing future voice acts).

Implications for workplace voice and listening literatures

Beyond safety, our findings have significance for both the workplace voice and listening literatures. For example, researchers often theorize listening in terms of attitudes and perceptions around voice, rather than how individuals react and respond to the content or intentions of a voice act. This is especially important for determining whether high-consequence voice acts, like whistleblowing (e.g., rather than small talk), are listened to.

Similarly, the voice literature generally assumes that listening will occur following speaking-up; accordingly, responses to such voice are rarely examined. Yet, voice and listening are iterative and complementary speech acts. Oftentimes safety or ethical problems are not resolved through single voice acts, but a series of voice acts and responses (e.g., during problem-solving), with some focusing on addressing and understanding concerns (e.g., potential medical errors), and others on ensuring third-turn repair (Schegloff, 1992). As such, listening may not always occur after a single voice act, but a sequence of voice acts that, potentially, have inconsistent information yet demand attention (Macnamara, 2018). In such exchanges, boundaries between voicers and listeners may become blurred, as listeners may transition to being voicers in team sense-making processes and while clarifying concerns.

Likewise, the role of who "voices" and who "listens" also requires consideration. For example, listeners can become future voicers if they cannot address voicers' concerns directly and must voice upwards. Vandekerckhove and colleagues (2012, 2014) conceptualize the process of voicing upwards as follows: if, after hearing concerns, listeners do not believe

they can address the problem, they can voice the concern to someone who can. Thus, listeners can become future voicers, with the voice/listening sequence continuing until the problem is addressed or the concern is discarded. Such systems are limited in that listeners must have the courage to voice to higher ups (Vandekerckhove & Langenberg, 2012), and must accurately recall and relay concerns (Tiitinen, 2020). To avoid perceptions that complaints were left unheard—and to encourage future voice (King et al., 2019)—listeners should keep voicers informed by communicating decisions related to concerns back to voicers, including if concerns will not be addressed. Future empirical research should examine how safety voice cascades upwards in organizations and how listeners can voice upwards while keeping voicers abreast.

The voice literature should also explore how voice and listening occur in multi-member teams, in organizations with multiple reporting systems, and outside of organizations (e.g., to regulators). This suggestion is especially pertinent to safety research, where individuals often engage in undirected voice to multi-person audiences (e.g., nurses pointing out abnormal bleeding to surgery teams; Kolbe et al., 2014), and listeners may not always know who is expected to respond. Poorly demarcated internal reporting systems may lead voicers to mistakenly believe that they raised formal complaints despite using informal channels (van Dael et al., 2022); likewise, voicers may have unclear concerns and may voice to incorrect audiences. It would be helpful to understand what happens following such misunderstandings and how organizations might correct this.

Safety listening: relationship with climate and culture?

The safety literature conceptualizes voice as a component of both safety climate and safety culture, and includes voice-related survey items in climate/culture assessments (Bisbey et al., 2021). While some organizational safety climate scales include listening-related items (e.g., Bahari & Clarke, 2013; Huang et al., 2013), current organizational culture conceptualizations typically do not include safety listening, despite its essential role in nurturing safe and ethical cultures and preventing potential scandals. Like Hald et al. (2020), we propose that safety listening should be seen as a subset of organizational cultures, warranting its inclusion in assessments. For instance, scholars may incorporate safety listening items into pre-existing culture surveys and capture safety listening through unobtrusive indicators of culture (e.g., quality of responses to whistleblowing complaints). This integration would provide a more comprehensive understanding of how speaking-up and listening behaviors combine to contribute to organizational cultures.

New research avenues

Here, we outline conceptual, theoretical, methodological, and empirical gaps, and recommend avenues for future research.

Proposing a standard term/definition. The review highlights a duplication of listening terms/definitions and a predominantly motivational focus, which likely impedes knowledge accumulation and theory development. It remains unclear whether the 36 terms are interchangeable or represent distinct concepts—for example, would improvable responses be opposite to positive or neutral ones? Research streams appear to be advancing independently, potentially leaving gaps in comprehending the spectrum of listening behaviors and their underlying drivers.

Expanding existing terms/definitions, we propose a standardized concept—safety listening—defined as listeners' behavior responding to safety voice demanding action to prevent harms in organizational contexts. Our definition emphasizes listeners' observable responses to voice acts (e.g., engaging, ignoring) as they directly influence outcomes (e.g., hazard mitigation). We underscore that voice acts may be

inaccurate (e.g., voicers may have incomplete information), therefore inaction may be appropriate following erroneous concerns. Crucially, effective listening is not necessarily agreeing with voicers; rather, it includes actions like investigation, intervention, and inaction (e.g., for false alarms). Our safety listening conceptualization encompasses both promotive and prohibitive voice messages—for instance, when individuals voice about improving safety (e.g., requesting an improved ventilation system), there is likely still an underlying concern (i.e., better ventilation would reduce the likelihood of harm). Thus, raising safety-related suggestions and concerns both include sharing observations or safety-related information to improve the status quo.

Extending explanations beyond listeners' motivations. Scholars typically position safety listening as resulting from listeners' motivations. This conceptualization likely influences and is influenced by challenges obtaining naturalistic listening data. Difficulties with reliably capturing listening behavior likely led to using self-report methodologies; these methodologies' findings are likely interpreted in terms of listeners' motivations due to biases and misattributions (e.g., voicers may attribute poor listening as motivational due to the self-attribution error). Moreover, conceptualizations framing listening as motivational indicate the use of self-report data assuming that listening is deliberate; these data then provide supporting evidence for explaining listening in terms of motivations. Thus, conceptualizations of ineffective listening and reliance upon self-report measures likely mutually reinforce each other and have constrained our understandings of listening.

Research should move beyond the focus on motivation-driven conceptualizations, especially since organizational disasters reveal instances where listeners were motivated to listen but failed to do so (e.g., pilots are motivated to safely fly airplanes; Noort et al., 2021a). Consequently, Martin et al. (2021) have questioned the prevailing notion that

mishandling complaints solely arises from deliberate efforts to enforce silence in organizations. We concur—recognizing that while motivations undoubtedly influence listening behaviors, concentrating solely on listeners' attitudes narrows our understandings and neglects alternative explanations.

The literature would benefit from developing a conceptual model and theory, which explain the organizational and system dynamics through which safety voice and listening impact outcomes (e.g., scandals). A plausible explanation posits that safety voice signifies shared cognition discrepancies within teams, while safety listening serves to re-establish shared cognition. Empirically investigating this proposition in future research may hold promise and researchers should continue refining explanations for the relationships between voice, listening, and outcomes.

Assessing safety listening behaviors. Most empirical studies employed self-report measures rather than assessing naturalistic listening behavior. Using proxies is understandable because safety listening behavior is elusive and requires infrequent complaints to occur. Moreover, truly high consequence experiments and simulations are unethical as they expose participants to significant risks.

Yet, relying on self-report, simulation, and experimental methodologies assumes that findings in safe and controlled situations generalize to dynamic and dangerous environments (Diener et al., 2022). In such decontextualized conditions, listening intentions or behaviors may be over-reported or inaccurately described, recalled, or attributed. For instance, self-report methods assume that individuals can (and would) precisely describe and attribute intentions to their own and others' listening behaviors. Human errors (e.g., misunderstandings), social desirability biases (van de Mortel, 2008), attribution errors (Ross, 1977), and other factors (e.g., primes; Bargh & Chartrand, 2000) may influence listening behaviors. Similarly, vignettes measuring listening

intentions may be abstract, over-rely on participants' imaginations, and may inaccurately predict listening behavior (Sheeran, 2005). Moreover, simulations and experiments involving confederates require precise execution of researchers' instructions to be believed by participants (Yeomans et al., 2023). Addressing discrepancies between self-reported perspectives (e.g., listener says they listened; voicer disagrees) poses another challenge. Such discrepancies may be frequent, as Bodie et al. (2014) found no association between voicers' perceptions of receivers' listening, receivers' perceptions of their listening, and behavioral listening measures. Likewise, voicers may incorrectly believe that listeners can address problems; however, listeners may hear safety concerns but cannot transparently show what action has been taken (e.g., addressing complaints about another's errors) and cannot do more than pass the complaint onwards.

Although self-report measures are helpful for addressing specific research inquiries (e.g., assessing employees' commitment to safety), they offer limited insights into the actual behaviors and underlying mechanisms driving individuals' responses to safety voice. Self-reports' limitations are heightened in this context because this measurement form is truly not high consequence in nature.

Only two out of 46 studies assessed listening behavior. Like Baumeister et al. (2007), we advocate for a more balanced approach in future research, with more behavioral measures using naturalistic data. We describe methodologies to measure naturalistic safety listening behaviors—including naturalistic and ecological observations and behavioral trace data—in Table 7. Measuring naturalistic listening behaviors would enable us to validate assumptions about this phenomenon (e.g., is effective safety listening always agreement?), uncover unexpected "real world" manifestations (e.g., listening to conflicting voice messages), and causally examine safety listening's relationship with important outcomes (e.g., airplane crashes). Additionally, researchers can triangulate behavioral and non-behavioral findings to converge on evidence that is more compelling than generated by one method alone (Barnes et al., 2018). Table 8 illustrates sample high validity behavioral trace datasets, which are publicly available and are unobtrusive (Hill et al., 2014). Due to its high validity, such data may be distressing; scholars should consider how to safeguard participants, researchers, and transcribers during data collection and analysis.

There is little consensus on how to code safety listening behaviors, likely due to an unclear understanding of what such behaviors entail in naturalistic settings. Studies, often using self-reported data, may have conflated effective listening with agreement; however, effective listening may be disagreement if voicers are incorrect. To address this, researchers should establish a taxonomy of safety listening behaviors, which can be translated into a reliable and valid coding framework. This framework should be empirically grounded and incorporate observable listening behaviors (e.g., asking questions; Kluger & Izchakov, 2022) and defensive tactics (Gillespie, 2020).

Given the considerable sample size of some of the behavioral trace datasets in Table 8 (i.e., thousands of instances of safety voice and safety listening), machine-learning-based natural language processing models and more recently large language models may revolutionize safety listening measurement. These techniques can rapidly analyze vast amounts of complex textual communications, generating high-quality results approaching human-level performance (Kjell et al., 2019; Luo et al., 2024; Törnberg, 2023). Accordingly, these techniques could efficiently and accurately identify patterns, trends, and potential risks in transcribed or written safety conversations and may detect novel nuanced insights that might not be possible with manual coders (Berger & Packard, 2022; Speer et al., 2022). For instance, artificial intelligence measures of safety voice and listening could code data at a sufficient scale, rigor, and subtlety to begin to identify

Table 7. Methodologies for Assessing Naturalistic Safety Listening Behaviors

Methodology	Description	Example	General benefits	General drawbacks
Naturalistic	observation	Researcher(s) observe dyad or team safety voice and listening behaviors within organizations	Lemke et al. (2021) observed safety voice and listening in healthcare teams as they were administering anesthesia pre-surgery, coding behaviors *in situ*	Can be used to empirically test theories Can be supplemented with interviews or follow-up questions to explore thinking behind actions
Challenges obtaining organizational access	Time-consuming data collection Individuals' behavior changing during observations Typically small sample sizes			
Ecological	observation	Participants wear portable recorders (e.g., body cameras; Electronically Activated Recorder) which record video and/or sounds as they go about their day (Mehl, 2017)	To our knowledge, has not yet been used to capture safety listening. A hypothetical example would be if Freedom to Speak Up Guardians (hospital employees designated to triage healthcare concerns) wore Electronically Activated Recorders	High quality recordings can capture subtle behaviors (e.g., sighs) Can be used to empirically test theories Minimally bothersome for participants Can be supplemented with interviews to explore thinking behind actions
Devices can be expensive Susceptible to technology issues Privacy	considerations Individuals' behavior changing during observations			
Behavioral trace data	Transcripts and/or recordings of real-life	Noort et al. (2021a) coded safety voice and listening in 172	Can be used to empirically test theories	Freedom of information

Table 7. (continued)

Methodology	Description	Example	General benefits	General drawbacks
	conversations containing safety voice and listening	cockpit conversation recordings from crashed aircraft	Unobtrusive Free, publicly available data lends to research reproducibility and "levels the playing field" for researchers in institutions with limited resources Sample sizes can be large (see Table 8)	requests may take time Data may be missing Recordings may be low-quality Challenging to explore decision-making behind behaviors

the types of voice and listening behaviors, which are associated with aviation accidents.

Proposing novel safety listening antecedents. Here, we suggest additional possible cognitive, interactional, and structural factors influencing safety listening.

Cognitive/skill-based factors

Stress likely influences safety listening, yet this relationship and its underlying mechanisms have not been empirically examined. Studies have shown that listening to trauma increases listeners' stress (Michelson & Kluger, 2021) and clinicians reported that stress influenced their listening (Long et al., 2020). We propose that stress impairs listening by diminishing cognitive capabilities and information processing (Sandi, 2013), which are required for situation awareness and listening abilities. Future research should examine how stressful environments influence safety listening and the skills required to effectively listen under stress.

Safety listening can be viewed as a skill rather than just an attitude, aligning with crew resource management (Kanki et al., 2019), non-technical skills (Fletcher et al., 2004), and workplace listening literatures (Itzchakov, 2020). This perspective is supported by listeners reporting they were under-skilled in

responding to complaints (Barlow et al., 2023a). In addition to general listening skills like suspending judgment (Itzchakov, 2020), safety listening skills encompass recognizing voice, determining when and what to listen to, and listening under stress. For instance, listeners must notice muted voice (Noort et al., 2021b) and undirected concerns (Kolbe et al., 2014), assess concerns' legitimacy, and discern between conflicting voice messages. These skills are trainable, as evidenced by Noort et al. (2021a), who found improved listening following crew resource management training. Training programs could incorporate real conversation recordings to help participants assess how others in their roles communicate (Stokoe, 2014).

Interactional factors

Groupthink—where the group's urge for conformity impedes the critical evaluation of signals indicating problems (Janis, 1972)—may discourage effective safety listening. Groupthink symptoms which can obstruct listening include collectively rationalizing warnings, categorizing voicers as inferior, and enforcing conformity by pressuring dissenters (Mannion & Thompson, 2014).

Interactions with technology and environments can impede the reception of voice, a factor often overlooked in the literature.

Table 8. Example Behavioral Trace Data Sources to Study Safety Listening

Data	Data source	Description
Aviation conversations	Noort et al., (2021c)	172 transcripts of conversations preceding aircraft crashes worldwide from 1962 to 2018
	VASAviation (2023)	Approximately 1,000 audio recordings of radio communication between cockpit crews and air traffic control preceding aircraft crashes and near misses worldwide from 2015 onwards
	Federal Aviation Administration (2019)	Approximately 200 audio recordings of radio communication between cockpit crews and air traffic control preceding aircraft crashes and near misses from 2006 to 2019 in the United States
Space mission conversations	National Aeronautics and Space Administration (2010)	80 transcripts of conversations between astronauts and ground control in the Mercury, Gemini, and Apollo space missions
Responses to healthcare feedback	Care Opinion United Kingdom (2023)	Approximately 14,000 healthcare staff responses to online comments from patients, families, and staff to United Kingdom healthcare institutions
	Care Opinion Australia (2023)	Approximately 2,700 healthcare staff responses to online comments from patients, families, and staff to Australian healthcare institutions
	Care Opinion Ireland (2023)	Approximately 100 healthcare staff responses to online comments from patients, families, and staff to Irish healthcare institutions
Enron email corpus	Jamison & Gurevych (n.d.)	Approximately 500,000 emails from 150 users (mostly Enron senior management) which the Federal Energy Regulatory Commission made public during its investigation of Enron for accounting fraud
Flint water crisis emails	Nowling & Seeger (2020)	37,376 pages of e-mails and related documents sent by or to the Executive Office of the Governor or the Michigan Department of Environmental Quality regarding Flint's drinking water quality. These were made public during the crisis' investigation
COVID-19 emails	Anthony Fauci Emails (n.d.)	Approximately 3,250 pages of emails received and responded to by Dr. Anthony Fauci during the COVID-19 pandemic which were made public following freedom of information requests
Boeing 737 MAX emails	Boeing 737 MAX Emails (n.d.)	Approximately 120 pages of emails received and responded to by Boeing staff regarding the faulty 737 MAX aircraft. These were publicly released by Boeing

Wilson et al. (2007) illustrate that faulty communication devices (e.g., dead batteries), human errors (e.g., mismatched radio frequencies), background noise (e.g., gunfire), and environmental obstacles (e.g., terrain obstructing radio signals) can impede voice's reception. Likewise, concerns may be unheard (e.g., whispered), incorrectly sent (e.g., wrong address), or

lost (e.g., improperly archived). Future research may consider hearing as a mediator between safety voice and safety listening and explore this relationship empirically.

Given that many communications rely on technology, both verbal (e.g., phone calls) and written (e.g., email, instant messaging), future research should explore whether safety voice and listening using these channels differ from face-to-face communication. Some behavioral trace data in Table 8 may aid this endeavor. It may be that non-verbal information gleaned from face-to-face or video communication (e.g., nodding) better facilitates developing shared understandings. It could also be that—with avenues to publicly complain about organizational practices (e.g., Glassdoor, blogs)—having a written and public record of concerns may prompt listeners' action to address hazards to "save face". Moreover, certain technology-mediated communications (e.g., phone calls) prompt instant safety listening while others (e.g., emails) may not be immediately responded to.

Structural factors

Organizational policies and procedures aimed at addressing specific wrongdoings have received limited attention in the literature despite their potential impact on safety listening. Organizational failure investigations reveal that these policies were either absent, unclear, or insufficient in guiding complaint handling (e.g., Crofts, 2017). These investigations also underscore that protocols can fail in unforeseen circumstances. For instance, in response to the September 11, 2001 attacks, organizations initially followed standard hijack protocols assuming the hijackers would make demands upon landing; however, this protocol was deemed inadequate for hijackings as attacks (Waller & Uitdewilligen, 2008). Consequently, we propose that clear and adequate policies/procedures highlighting their possible fallibility in unexpected situations would encourage safety listening.

Conclusion

Research has highlighted the significance of voicing high consequence concerns to avert harm. Nonetheless, although voicing is often necessary, it is insufficient as evident in many organizational disasters where raised concerns went unaddressed. Recognizing this, we conducted this integrated conceptual review to establish the concept of safety listening as the necessary counterpart to safety voice. This review synthesizes existing publications to define safety listening as listeners' behavioral responses to safety voice acts in organizational settings which are intended to avoid physical and/or social harms. In advancing the field, we distinguish safety listening from other listening forms, recommend non-motivational explanations, advocate for the utilization of naturalistic data to measure listening behaviors, and suggest novel contributory factors. This review lays the foundation for future research to foster a comprehensive and cumulative understanding of safety listening, ultimately contributing to the prevention of future organizational failures.

ORCID iD

Alyssa M. Pandolfo [iD] https://orcid.org/0000-0003-0455-496X

References

Anthony Fauci Emails. (n.d.). Retrieved March 7, 2023, from https://www.documentcloud.org/documents/20793561-leopoldnih-foia-anthony-fauci-emails

Atkinson, C. D. E., Wesselmann, E. D. & Lannin, D. G. (2022). Understanding why some whistle-blowers are venerated and others vilified. *Heroism Science*, 7(2), 1–34. https://doi.org/10.26736/hs.2022.01.06

Bahari, S. F. & Clarke, S. (2013). Cross-validation of an employee safety climate model in Malaysia. *Journal of Safety Research*, 45, 1–6. https://doi.org/10.1016/j.jsr.2012.12.003

Bargh, J. A. & Chartrand, T. L. (2000). The mind in the middle: A practical guide to priming and automaticity research. In *Handbook of research methods in social and personality psychology* (pp. 253–285). Cambridge University Press. https://doi.org/10.1163/_afco_asc_1327.

Barlow, M. (2021). Enhancing intergroup communication in healthcare: The role of the receiver. *Focus on Health Professional Education: A Multi-Professional Journal*, 22(3), 78–84. https://doi.org/10.11157/fohpe.v22i3.523

Barlow, M., Morse, K. J., Watson, B. & Maccallum, F. (2023a). Identification of the barriers and enablers for receiving a speaking up message: A content analysis approach. *Advances in Simulation*, 8(17), 1–12. https://doi.org/10.1186/s41077-023-00256-1

Barlow, M., Watson, B. & Jones, E. (2023b). Understanding observed receiver strategies in the healthcare speaking up context. *International Journal of Healthcare Simulation*, *Xx*, 1–12. https://doi.org/10.54531/SUFD5615

Barlow, M., Watson, B. & Rudolph, J. (2019). How being a great receiver can change the game in speaking up conversations. *MedEdPublish*, 8(3), 1–9. https://doi.org/10.15694/mep.2019.000204.1

Barnes, C. M., Dang, C. T., Leavitt, K., Guarana, C. L. & Uhlmann, E. L. (2018). Archival data in micro-organizational research: A toolkit for moving to a broader set of topics. *Journal of Management*, 44(4), 1453–1478. https://doi.org/10.1177/0149206315604188

Baumeister, R. F., Vohs, K. D. & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, 2(4), 396–403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Bazzoli, A. & Curcuruto, M. (2021). Safety leadership and safety voices: Exploring the mediation role of proactive motivations. *Journal of Risk Research*, 24(11), 1368–1387. https://doi.org/10.1080/13669877.2020.1863846

BBC News. (2018). Larry Nassar case: USA Gymnastics doctor "abused 265 girls".

Berger, J. & Packard, G. (2022). Using natural language processing to understand people and culture. *American Psychologist*, 77(4), 525–537. https://doi.org/10.1037/amp0000882

Bisbey, T. M., Kilcullen, M. P., Thomas, E. J., Ottosen, M. J., Tsao, K. J. & Salas, E. (2021). Safety culture: An integration of existing models and a framework for understanding its development. *Human Factors*, 63(1), 88–110. https://doi.org/10.1177/0018720819868878

Blenkinsopp, J., Snowden, N., Manion, R., Powell, M., Davies, H. T. O., Millar, R. & McHale, J. (2019). Whistleblowing over patient safety and care quality: A review of the literature. *Journal of Health Organization and Management*, 33(6), 737–756. https://doi.org/10.1108/JHOM-12-2018-0363

Boeing 737 MAX Emails. (n.d.). https://www.documentcloud.org/documents/6615980-Boeing-737-MAX-emails

Bodie, G. D., Jones, S. M., Vickery, A. J., Hatcher, L. & Cannava, K. (2014). Examining the construct validity of enacted support: A multitrait–multimethod analysis of three perspectives for judging immediacy and

listening behaviors. *Communication Monographs*, *81*(4), 495–523. https://doi.org/10.1080/03637751.2014.957223

Brown, M. E. & Treviño, L. K. (2006). Ethical leadership: A review and future directions. *Leadership Quarterly*, *17*(6), 595–616. https://doi.org/10.1016/j.leaqua.2006.10.004

Care Opinion Australia. (2023). https://www.careopinion.org.au/

Care Opinion Ireland. (2023). https://www.careopinion.ie/

Care Opinion United Kingdom. (2023). https://www.careopinion.org.uk/opinions

Casal, J. C. & Zalkind, S. S. (1995). Consequences of whistle-blowing: A study of the experiences of management accountants. *Psychological Reports*, *77*(3), 795–802. https://doi.org/10.2466/pr0.1995.77.3.795.

Chamberlin, M., Newton, D. W. & Lepine, J. A. (2017). A meta-analysis of voice and its promotive and prohibitive forms: Identification of key associations, distinctions, and future research directions. *Personnel Psychology*, *70*(1), 11–71. https://doi.org/10.1111/peps.12185

Chen, A. & Treviño, L. K. (2022). Promotive and prohibitive ethical voice: Coworker emotions and support for the voice. *Journal of Applied Psychology*, *107*(11), 1–22. https://doi.org/10.1037/apl0001003.

Chen, A. & Treviño, L. K. (2023). The consequences of ethical voice inside the organization: An integrative review. *Journal of Applied Psychology*, *108*(8), 1316–1335. https://doi.org/10.1037/apl0001075

Cleary, S. R. & Duke, M. (2019). Clinical governance breakdown: Australian cases of wilful blindness and whistleblowing. *Nursing Ethics*, *26*(4), 1039–1049. https://doi.org/10.1177/0969733017731917

CNN Editorial Research. (2021). *Flint water crisis fast facts*. CNN. https://edition.cnn.com/2016/03/04/us/flint-water-crisisfast-facts/index.html

Cortina, L. M. & Magley, V. J. (2003). Raising voice, risking retaliation: Events following interpersonal mistreatment in the workplace. *Journal of Occupational Health Psychology*, *8*(4), 247–265. https://doi.org/10.1037/1076-8998.8.4.247

Crofts, P. (2017). Criminalising institutional failures to prevent, identify or react to child sexual abuse. *International Journal for Crime, Justice and Social Democracy*, *6*(3), 104–122. https://doi.org/10.5204/ijcjsd.v6i3.421

Cuellar, M. J., Keil, M. & Johnson, R. D. (2006). The deaf effect response to bad news reporting in information systems projects. *E-Service Journal*, *5*(1), 75. https://doi.org/10.2979/esj.2006.5.1.75

Dator, J. (2021). *A comprehensive timeline of the Larry Nassar case*. SBNation. https://www.sbnation.com/2018/1/19/16900674/larry-nassar-abuse-timeline-usa-gymnasticsmichigan-state

Diener, E., Northcott, R., Zyphur, M. J. & West, S. G. (2022). Beyond experiments. *Perspectives on Psychological Science*, *17*(4), 1101–1119. https://doi.org/10.1177/17456916211037670

Edmondson, A. C. (2018). *The fearless organization: Creating psychological safety in the workplace for learning, innovation, and growth*. John Wiley & Sons.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Federal Aviation Administration. (2019). Air Traffic Control audio. Accident & Incident Data. https://www.faa.gov/data_research/accident_incident

Fernando, D. & Prasad, A. (2019). Sex-based harassment and organizational silencing: How women are led to reluctant acquiescence in academia. *Human Relations*, *72*(10), 1565–1594. https://doi.org/10.1177/0018726718809164.

Fletcher, G., Flin, R., McGeorge, P., Glavin, R., Maran, N. & Patey, R. (2004). Rating non-technical skills: Developing a behavioural marker system for use in anaesthesia. *Cognition, Technology and Work*, *6*, 165–171. https://doi.org/10.1007/s10111-004-0158-y

Gillespie, A. (2020). Disruption, self-presentation, and defensive tactics at the threshold of learning. *Review of General Psychology*, *24*(4), 382–396. https://doi.org/10.1177/1089268020914258

Groves, P. S., Bunch, J. L., Cannava, K. E., Sabadosa, K. A. & Williams, J. K. (2021). Nurse sensemaking for responding to patient and family safety concerns. *Nursing Research*, *70*(2), 106–113. https://doi.org/10.1097/NNR.0000000000000487

Hald, E. J., Gillespie, A. & Reader, T. W. (2020). Causal and corrective organisational culture: A systematic review of case studies of institutional failure. *Journal of Business Ethics*. 174, 457–483. https://doi.org/10.1007/s10551-020-04620-3.

Hällgren, M., Rouleau, L. & De Rond, M. (2018). A matter of life or death: How extreme context research matters for management and organization studies. *Academy of Management Annals*, *12*(1), 111–153. https://doi.org/10.5465/annals.2016.0017

Harlos, K. P. (2001). When organizational voice systems fail: More on the deaf-ear syndrome and frustration effects. *The Journal of Applied Behavioral Science*, *37*(3), 324–342. https://doi.org/10.1177/0021886301373005

Herrnstein, R. J. & Prelec, D. (1991). Melioration: A theory of distributed choice. *Journal of Economic Perspectives*, *5*(3), 137–156. https://doi.org/10.1257/jep.5.3.137

Hill, A. D., White, M. A. & Wallace, J. C. (2014). Unobtrusive measurement of psychological constructs in organizational research. *Organizational Psychology Review*, *4*(2), 148–174. https://doi.org/10.1177/2041386613505613

Hsieh, S. Y., Thomas, D. & Rotem, A. (2005). The organisational response to patient complaints: A case study in Taiwan. *International Journal of Health Care Quality Assurance*, *18*(4), 308–320. https://doi.org/10.1108/09526860510602578

Huang, Y. H., Zohar, D., Robertson, M. M., Garabet, A., Lee, J. & Murphy, L. A. (2013). Development and validation of safety climate scales for lone workers using truck drivers as exemplar. *Transportation Research Part F: Traffic Psychology and Behaviour*, *17*, 5–19. https://doi.org/10.1016/j.trf.2012.08.011

Hulland, J. (2020). Conceptual review papers: Revisiting existing research to develop and refine theory. *AMS Review*, *10*(1–2), 27–35. https://doi.org/10.1007/s13162-020-00168-7

Itzchakov, G. (2020). Can listening training empower service employees? The mediating roles of anxiety and perspective-taking. *European Journal of Work and Organizational Psychology*, *29*(6), 938–952. https://doi.org/10.1080/1359432X.2020.1776701

Itzchakov, G. & Kluger, A. N. (2017). Can holding a stick improve listening at work? The effect of Listening Circles on employees' emotions and cognitions. *European Journal of Work and Organizational Psychology*, *26*(5), 663–676. https://doi.org/10.1080/1359432X.2017.1351429

Jamison, E. K. & Gurevych, I. (n.d.). Enron threads corpus and Enron crowdsourced dataset. Retrieved July 10, 2022, from https://tudatalib.ulb.tu-darmstadt.de/handle/tudatalib/2456

Janis, I. (1972). *Victims of groupthink*.

Jones, A., Blake, J., Adams, M., Kelly, D., Mannion, R. & Maben, J. (2021). Interventions promoting employee "speaking-up" within healthcare workplaces: A systematic narrative review of the international literature. *Health Policy*, *125*(3), 375–384. https://doi.org/10.1016/j.healthpol.2020.12.016

Jones, A. & Kelly, D. (2014). Deafening silence? Time to reconsider whether organisations are silent or deaf when things go wrong. *BMJ Quality and Safety*, *23*(9), 709–713. https://doi.org/10.1136/bmjqs-2013-002718

Kanki, B. G., Anca, J. & Chidester, T. R. (eds) (2019). *Crew resource management* (3rd ed.). Academic Press.

Kenny, K. (2019). *Whistleblowing: Toward a new theory*. Harvard University Press.

Kenny, K., Fotaki, M. & Scriver, S. (2019). Mental health as a weapon: Whistleblower retaliation and normative violence. *Journal of Business Ethics*, *160*(3), 801–815. https://doi.org/10.1007/s10551-018-3868-4

King, D. D., Ryan, A. M. & Van Dyne, L. (2019). Fostering future voice after non-endorsement of suggestions. *Journal of Occupational and Organizational Psychology*, *92*, 535–565. https://doi.org/10.1111/joop.12275

Kirby, J. (2018, May 16). The sex abuse scandal surrounding USA Gymnastics team doctor Larry Nassar, explained. *Vox.Com*. https://www.vox.com/identities/2018/1/19/16897722/sexual-abuse-usa-gymnastics-larry-nassar-explained

Kjell, O. N. E., Kjell, K., Garcia, D. & Sikstrom, S. (2019). Semantic measures: Using natural language processing to measure, differentiate, and describe psychological constructs. *Psychological Methods*, *24*(1), 92–115. https://doi.org/10.1037/met0000191.supp

Kluger, A. N. & Izchakov, G. (2022). The power of listening at work. *Annual Review of Organizational Psychology and Organizational Behavior*, *7*(01), 46–54. https://doi.org/10.1146/annurev-orgpsych-012420-091013

Kluger, A. N., Lehmann, M., Aguinis, H., Itzchakov, G., Gordoni, G., Zyberaj, J. & Bakaç, C. (2024). A meta-analytic systematic review and theory of the effects of perceived listening on work outcomes. *Journal of Business and Psychology*. *39*(2), 295–344. https://doi.org/10.1007/s10869-023-09897-5.

Kolbe, M., Grote, G., Waller, M. J., Wacker, J., Grande, B., Burtscher, M. J. & Spahn, D. R. (2014). Monitoring and talking to the room: Autochthonous coordination patterns in team interaction and performance. *Journal of Applied Psychology*, *99*(6), 1254–1267. https://doi.org/10.1037/a0037877

Krenz, H. L. & Burtscher, M. J. (2021). Investigating voice in action teams: A critical review. *Cognition, Technology and Work*, *23*(3), 605–624. https://doi.org/10.1007/s10111-020-00646-9

Kundro, T. G. & Rothbard, N. P. (2023). Does power protect female moral objectors? How and when moral objectors' gender, power, and use of organizational frames influence perceived self-control and experienced retaliation. *Academy of Management Journal*, *66*(1), 306–334.

Lainidi, O., Jendeby, M. K., Montgomery, A., Mouratidis, C., Paitaridou, K., Cook, C., Johnson, J. & Karakasidou, E. (2023). An integrative systematic review of employee silence and voice in healthcare: What are we really measuring? *Frontiers in Psychiatry*, *14*(May), 1–28. https://doi.org/10.3389/fpsyt.2023.1111579

Lemke, R., Burtscher, M. J., Seelandt, J. C., Grande, B. & Kolbe, M. (2021). Associations of form and function of speaking up in anaesthesia: A prospective observational study. *British Journal of Anaesthesia*, *127*(6), 971–980. https://doi.org/10.1016/j.bja.2021.08.014.

Long, J., Jowsey, T., Garden, A., Henderson, K. & Weller, J. (2020). The flip side of speaking up: A new model to facilitate positive responses to speaking up in the operating theatre. *British Journal of Anaesthesia*, *125*(6), 1099–1106. https://doi.org/10.1016/j.bja.2020.08.025

Luo, X., Rechardt, A., Sun, G., Nejad, K. K., Yáñez, F., Yilmaz, B., Lee, K., Cohen, A. O., Borghesani, V., Pashkov, A., Marinazzo, D., Nicholas, J., Salatiello, A., Sucholutsky, I., Minervini, P., Razavi, S., Rocca, R., Yusifov, E., Okalova, T. & Love, B. C. (2024). *Large language models surpass human experts in predicting neuroscience results*. http://arxiv.org/abs/2403.03230.

Mack, J. & Lawler, E. (2017). MSU doctor's alleged victims talked for 20 years. Was anyone listening? Michigan Live. https://www.mlive.com/news/page/msu_doctor_alleged_sexual_assault.html

Macnamara, J. (2018). Toward a theory and practice of organizational listening. *International Journal of Listening*, *32*(1), 1–23. https://doi.org/10.1080/10904018.2017.1375076

Mannion, R. & Thompson, C. (2014). Systematic biases in group decision-making: Implications for patient safety. *International Journal for Quality in Health Care*, *26*(6), 606–612. https://doi.org/10.1093/intqhc/mzu083

Martin, B. & Rifkin, W. (2004). The dynamics of employee dissent: Whistleblowers and organizational jiu-jitsu. *Public Organization Review*, *4*(3), 221–238. https://doi.org/10.1023/b:porj.0000036869.45076.39

Martin, G. P., Chew, S. & Dixon-Woods, M. (2021). Why do systems for responding to concerns and complaints so often fail patients, families and healthcare staff? A qualitative study. *Social Science and Medicine*, *287*, 114375. https://doi.org/10.1016/j.socscimed.2021.114375

McDonald, S. & Ahern, K. (2000). The professional consequences of whistleblowing by nurses. *Journal of Professional Nursing*, *16*(6), 313–321. https://doi.org/10.1053/jpnu.2000.18178

Mehl, M. R. (2017). The electronically activated recorder (EAR) a method for the naturalistic observation of daily social behavior. *Current Directions in Psychological Science*, *26*(2), 184–190.

Mesmer-Magnus, J. R. & Viswesvaran, C. (2005). Whistleblowing in organizations: An examination of correlates of whistleblowing intentions, actions, and retaliation. *Journal of Business Ethics*, *62*(3), 277–297. https://doi.org/10.1007/s10551-005-0849-1

Miceli, M. P., Near, J. P. & Dworkin, T. M. (2008). *Whistle-blowing in organizations*. Psychology Press. https://doi.org/10.4324/9780203809495

Miceli, M. P., Rehg, M., Near, J. P. & Ryan, K. C. (1999). Can laws protect whistle-blowers? Results of a naturally occurring field experiment. *Work and Occupations*, *26*(1), 129–151. https://doi.org/10.1177/0730888499026001007

Michelson, T. & Kluger, A. (2021). Can listening hurt you? A meta-analysis of the effects of exposure to trauma on listener's stress. *International Journal of Listening*, 1–11. https://doi.org/10.1080/10904018.2021.1927734

Morrison, E. W. (2011). Employee voice behavior: Integration and directions for future research. *Academy of Management Annals*, *5*(1), 373–412. https://doi.org/10.1080/19416520.2011.574506

Morrison, E. W. (2014). Employee voice and silence. *Annual Review of Organizational Psychology and Organizational Behavior*, *1*(1), 173–197. https://doi.org/10.1146/annurev-orgpsych-031413-091328

National Aeronautics and Space Administration. (2010). Communications transcripts: *Mercury through Apollo*. https://historycollection.jsc.nasa.gov/JSCHistoryPortal/history/mission_trans/mission_transcripts.htm

Near, J. P. & Jensen, T. C. (1983). The whistleblowing process: Retaliation and perceived effectiveness. *Work and Occupations*, *10*(1), 3–28. https://doi.org/10.1177/0730888483010001001

Near, J. P. & Miceli, M. P. (1985). Organizational dissidence: The case of whistle-blowing. *Journal of Business Ethics*, *4*, 1–16. https://doi.org/10.1007/978-94-007-4126-3_8

Near, J. P. & Miceli, M. P. (1986). Retaliation against whistle blowers: Predictors and effects. *Journal of Applied Psychology*, *71*(1), 137–145. https://doi.org/10.1037/0021-9010.71.1.137

Noort, M. C., Reader, T. W. & Gillespie, A. (2019). Speaking up to prevent harm: A systematic review of the safety voice literature. *Safety Science*, *117*(January), 375–387. https://doi.org/10.1016/j.ssci.2019.04.039

Noort, M. C., Reader, T. W. & Gillespie, A. (2021a). Safety voice and safety listening during aviation accidents: Cockpit voice recordings reveal that speaking-up to power is not enough. *Safety Science*, *139*(March), 105260. https://doi.org/10.1016/j.ssci.2021.105260

Noort, M. C., Reader, T. W. & Gillespie, A. (2021b). The sounds of safety silence: Interventions and temporal patterns unmute unique safety voice content in speech. *Safety Science*, *140*(June 2020), 105289. https://doi.org/10.1016/j.ssci.2021.105289

Noort, M. C., Reader, T. W. & Gillespie, A. (2021c). Cockpit voice recorder transcript data: Capturing safety voice and safety listening during historic aviation accidents. *Data in Brief*, *39*, 107602.

Nowling, W. D. & Seeger, M. W. (2020). Sensemaking and crisis revisited: The failure of sensemaking during the Flint water crisis. *Journal of Applied Communication Research*, *48*(2), 270–289. https://doi.org/10.1080/00909882.2020.1734224

Park, H., Bjørkelo, B. & Blenkinsopp, J. (2020). External whistleblowers' experiences of workplace bullying by superiors and colleagues. *Journal of Business Ethics*, *161*(3), 591–601. https://doi.org/10.1007/s10551-018-3936-9

Parmerlee, M. A., Near, J. P. & Jensen, T. C. (1982). Correlates of whistle-blowers' perceptions of organizational retaliation. *Administrative Science Quarterly*, *27*(1), 17–34.

Peirce, E., Smolinski, C. A. & Rosen, B. (1998). Why sexual harassment complaints fall on deaf ears. *Academy of Management Executive*, *12*(3), 41–54. https://doi.org/10.5465/ame.1998.1109049

Pierce, C. A., Aguinis, H. & Adams, S. K. (2000). Effects of a dissolved workplace romance and rater characteristics on responses to a sexual harassment accusation. *Academy of Management Journal*, *43*(5), 869–880.

Pierce, C. A., Broberg, B. J., McClure, J. R. & Aguinis, H. (2004). Responding to sexual harassment complaints: Effects of a dissolved workplace romance on decision-making standards. *Organizational Behavior and Human Decision Processes*, *95*(1), 66–82. https://doi.org/10.1016/j.obhdp.2003.06.001

Power, S. A., Zittoun, T., Akkerman, S., Wagoner, B., Cabra, M., Cornish, F., Hawlina, H., Heasman, B., Mahendran, K., Psaltis, C., Rajala, A., Veale, A. & Gillespie, A. (2023). Social psychology of and for world-making. *Personality and Social Psychology*

Review. *27*(4), 378–392. https://doi.org/10.1177/10888683221145756.

Reader, T. W. (2022). Stakeholder safety communication: Patient and family reports on safety risks in hospitals. *Journal of Risk Research*, *25*(7), 807–824. https://doi.org/10.1080/13669877.2022.2061036

Rehg, M. T., Miceli, M. P., Near, J. P. & Van Scotter, J. R. (2008). Antecedents and outcomes of retaliation against whistleblowers: Gender differences and power relationships. *Organization Science*, *19*(2), 221–240. https://doi.org/10.1287/orsc.1070.0310

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (10th ed., pp. 173–220). Academic Press.

Roulet, T. J. & Pichler, R. (2020). Blame game theory: Scapegoating, whistleblowing and discursive struggles following accusations of organizational misconduct. *Organization Theory*, *1(4)*, 2631787720975192. https://doi.org/10.1177/2631787720975192.

Salas, E., Bisbey, T. M., Traylor, A. M. & Rosen, M. A. (2020). Can teamwork promote safety in organizations? *Annual Review of Organizational Psychology and Organizational Behavior*, *7*, 283–313. https://doi.org/10.1146/annurev-orgpsych-012119-045411

Sandi, C. (2013). Stress and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(3), 245–261. https://doi.org/10.1002/wcs.1222

Schegloff, E. A. (1992). Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology*, *97*(5), 1295–1345. https://doi.org/10.1086/229903

Schramm, W. (1955). *How communication works*. University of Illinois Press.

Shannon, C. E. & Weaver, W. (1949). *The mathematical theory of communication*. University of Illinois Press.

Sheeran, P. (2005). Intention-behavior relations: A conceptual and empirical review. *European Review of Social Psychology*, *12*(1), 1–36. https://doi.org/10.1002/0470013478.ch1

Siddaway, A. P., Wood, A. M. & Hedges, L. V. (2019). How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology*, *70*, 747–770. https://doi.org/10.1146/annurev-psych-010418-102803

Skinner, B. F. (1963). Operant behavior. *American Psychologist*, *18*(8), 503–515. https://doi.org/10.1037/h0045185

Speer, A. B., Perrotta, J., Tenbrink, A. P., Wegmeyer, L. J., Delacruz, A. Y. & Bowker, J. (2022). Turning words into numbers: Assessing work attitudes using natural language processing. *Journal of Applied Psychology*, *108*(6), 1027–1045. https://doi.org/10.1037/apl0001061

Stokoe, E. (2014). The Conversation Analytic Role-play Method (CARM): A method for training communication skills as an alternative to simulated role-play. *Research on Language and Social Interaction*, *47*(3), 255–265. https://doi.org/10.1080/08351813.2014.925663

Sumanth, J. J., Mayer, D. M. & Kay, V. S. (2011). Why good guys finish last: The role of justification motives, cognition, and emotion in predicting retaliation against whistleblowers. *Organizational Psychology Review*, *1*(2), 165–184. https://doi.org/10.1177/2041386611398283

Taylor, A. (2014, December 2). Bhopal: The world's worst industrial disaster, 30 years later. *The Atlantic*. https://www.theatlantic.com/photo/2014/12/bhopal-the-worlds-worst-industrial-disaster-30-years-later/100864/.

Tiitinen, L. (2020). The power of silence: Silencing as a method of preventing whistleblowing. *European Journal of Social Work*, *23*(1), 68–79. https://doi.org/10.1080/13691457.2018.1460323

Törnberg, P. (2023). *ChatGPT-4 outperforms experts and crowd workers in annotating political twitter messages with zero-shot learning. XXX*(Xx), 1–5. http://arxiv.org/abs/2304.06588.

Tucker, S. & Turner, N. (2015). Sometimes it hurts when supervisors don't listen: The antecedents and consequences of safety voice among young workers. *Journal of Occupational Health Psychology*, *20*(1), 72–81. https://doi.org/10.1037/a0037756

van Dael, J., Reader, T., Gillespie, A., Freise, L., Darzi, A. & Mayer, E. (2022). Do national policies for complaint handling in English hospitals support quality improvement? Lessons from a

case study. *Journal of the Royal Society of Medicine*, *115*(10), 390–398. https://doi.org/10.1177/01410768221098247.

Vandekerckhove, W., Brown, A. J. & Tsahuridu, E. (2014). Managerial responsiveness to whistle-blowing: Expanding the research horizon. *In International Handbook on Whistleblowing Research*, 298–328. https://doi.org/10.4337/9781781006795.00021

Vandekerckhove, W. & Langenberg, S. (2012). Can we organize courage? Implications from Foucault's parrhesia. *Electronic Journal of Business Ethics and Organizational Studies*, *17*(2), 35–44.

van de Mortel, T. (2008). Faking it: Social desirability response bias in self-report research. *Australian Journal of Advanced Nursing*, *25*(4), 40–48.

VASAviation. (2023). https://www.youtube.com/c/VASAviation

Waller, M. J. & Uitdewilligen, S. (2008). Talking to the room: Collective sensemaking during crisis situations. In *Time in organizational research (pp. 208–225)*. Routledge.

Weick, K. E. (1990). The vulnerable system: An analysis of the Tenerife air disaster. *Journal of Management*, *16*(3), 571–593. https://doi.org/10.1177/014920639001600304

Weick, K. E. (2010). Reflections on enacted sensemaking in the Bhopal disaster. *Journal of Management Studies*, *47*(3), 537–550.

Wellman, N., Mayer, D. M., Ong, M. & DeRue, D. S. (2016). When are do-gooders treated badly? Legitimate power, role expectations, and reactions to moral objection in organizations. *Journal of Applied Psychology*, *10*(6), 793–814. https://doi.org/10.1037/apl0000094

Westrum, R. (2004). A typology of organisational cultures. *Quality and Safety in Health Care*, *13*(Suppl. 2), 22–27. https://doi.org/10.1136/qshc.2003.009522

Westrum, R. (2014). The study of information flow: A personal journey. *Safety Science*, *67*, 58–63. https://doi.org/10.1016/j.ssci.2014.01.009

Wilkinson, A., Townsend, K., Graham, T. & Muurlink, O. (2015). Fatal consequences: An analysis of the failed employee voice system at the Bundaberg Hospital. *Asia Pacific Journal of Human Resources*, *53*(3), 265–280. https://doi.org/10.1111/1744-7941.12061

Wilson, K. A., Salas, E., Priest, H. A. & Andrews, D. (2007). Errors in the heat of battle: Taking a closer look at shared cognition breakdowns through teamwork. *Human Factors*, *49*(2), 243–256. https://doi.org/10.1518/001872007X312478

Yeomans, M., Boland, F. K., Collins, H. K., Abi-Esber, N. & Wood Brooks, A. (2023). *A practical guide to conversation research: How to study what people say to each other*. 1–23.

Yip, J. & Fisher, C. M. (2022). Listening in organizations: A synthesis and future agenda. *Academy of Management Annals*, *16*(2), 657–679. https://doi.org/10.5465/annals.2020.0367

Author biographies

Alyssa M. Pandolfo is a doctoral candidate in psychological and behavioral science at the London School of Economics. Her expertise is in how high-risk organizations (e.g., healthcare, aviation) can understand and improve safety-critical behaviors like voicing and listening to concerns. More broadly, her research interests include organizational decision-making, safety and ethical culture, teamwork, and resilience. Her work has been published in *BMJ Quality & Safety*, the *Canadian Journal of Experimental Psychology*, and the *Journal of Antimicrobial Chemotherapy*.

Tom W. Reader is an associate professor of Organizational Psychology at the London School of Economics. Working in 'high-risk' domains (e.g., healthcare, aviation), his research focuses on how organizations can become more resilient. For example, by improving voicing on conduct problems, listening to concerns, teamwork and decision-making, safety culture, and systems for reporting and learning. His work has been published in journals such as the *Journal of Applied Psychology*, *Risk Analysis*, and the *Journal of Business Ethics*.

Alex Gillespie is a professor of Psychological and Behavioral Science at the London School of Economics. His research begins with recognizing that people and groups have different perspectives. Key questions include: How do people bridge these divergences of perspective and achieve their collective life? And, more pragmatically, what technologies can we develop to study and help coordinate these perspectives? Methodologically, he likes analyzing interesting sources of textual data using mixed methods. His work has been published in journals including *Milbank Quarterly*, *Review of General Psychology*, and *Personality and Social Psychology Review*.