The London School of Economics and Political Science

THE BEHAVIOURAL NATURE OF SAFETY VOICE:

ADVANCING CONCEPTS AND MEASURES TO ENABLE THE PREVENTION OF HARM

A thesis submitted to the London School of Economics and Political Science's

Department of Psychological and Behavioural Science

for the degree of Doctor of Philosophy

August 2020

DECLARATION

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it). The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that this thesis consists of 75,840 words (excluding references and appendices).

I consider the work submitted to be a complete thesis fit for examination.

I authorise that, if a degree is awarded, an electronic copy of my thesis will be deposited in LSE Theses Online (in accordance with the published deposit agreement) held by the British Library of Political and Economic Science and that, except as provided for in regulation 61 it will be made available for public reference.

I authorise the School to supply a copy of the abstract of my thesis for inclusion in any published list of theses offered for higher degrees in British universities or in any supplement thereto, or for consultation in any central file of abstracts of such theses.

Statement of co-authored work

I confirm that Articles 1-4 were jointly co-authored with Dr Reader and Dr Gillespie and I contributed 80% of this work. For all other sections I am the sole author.

Statement of use of third party for editorial help

I can confirm that my thesis (bar the articles) was copy edited for conventions of language, spelling and grammar by the Language Centre at the London School of Economics and Political science.

ABSTRACT

Background: The concept of 'safety voice' captures the extent to which individuals speak-up about safety. The behaviour is deemed important for preventing accidents, yet interventions are needed because people often fail to speak-up ('safety silence'), thus contributing to harmful outcomes across safety-critical domains. However, the concept remains disintegrated and grounded in limited evidence and methodologies. Thus, the utility of 'safety voice' for safety management remains unclear, prohibiting effective interventions. This thesis therefore aims to evaluate how the behavioural nature of safety voice may be optimally conceptualised, assessed and intervened on. **Approach:** Four articles presented a systematic literature review ($n = 48$ publications), twelve experimental studies ($n_{total} = 1,222$) and an analysis of Cockpit Voice Recorder (CVR) transcripts across 172 aviation accidents (1962-2018; $n = 14,128$ conversational turns). Article 1 synthesised evidence from across theoretical domains. Article 2 presented the first experimental paradigm for safety voice ('Walking the Plank') to address nine methodological challenges. Article 3 observed safety silence in the laboratory to establish and conceptualise how the behaviour manifests in relationship to safety voice and interventions. Article 4 captured safety voice during real-life safety accidents, and investigated how risk, safety listening, power distance and CRM training impact on safety voice. **Findings:** Safety voice is a distinct concept that is highly ecological and situated, and that is important for understanding how safety voice contributes to accidents. A methodological reliance on self-reports and post-hoc methodologies was identified and addressed through the Walking the Plank paradigm. Safety silence, identifiable through assessing safety concerns, was scalable based on the degree of safety voice speech, with interventions uniquely impacting on five safety themes and hazard stages. Safety voice was found to occur frequently during real accidents, with the developed Threat Mitigation Model underscoring that safety concerns, safety voice and safety listening all contribute to preventing harm.

PERSONAL STATEMENT

In the spirit of my alma maters (i.e., the London School of Economics and Utrecht University), I care about conducting rigorous research in order to create a better world and address real-world problems. Conducting research on safety voice fits this personal mission. This is because by investigating the nature of the safety voice I can contribute new insights into how people may be enabled to prevent harm. The problem posed by people withholding safety concerns is not merely academic but has a very clear application for understanding and improving situations in which people encounter hazards: these situations can have serious outcomes such as injuries, accidents, or organisational decline. Whilst I acknowledge that mishaps are a normal part of social and institutional life (Perrow, 2011), I assume that at least a portion of adverse events are preventable through policies and behaviours that mitigate contributing causes (Nabhan et al., 2012). This assumption is important to me because it provides me with a hope that a better world is possible, one in which people can manage risks effectively, and that knowledge may be developed to support this. This drives me to investigate ways for people and institutions to create better outcomes, and motivates me to become a better researcher. Of course, I do not claim that creating a better world through solving real problems is the only worthy endeavour: arguably, pursuing knowledge is virtuous in and of itself. Yet, this simple motivation (to apply fundamental knowledge in order to enable people to do good) and hope (that harmful outcomes are preventable) drives me to investigate phenomena such as safety voice.

Through investigating safety voice, I continue my line of work on the psychology of risk within social and organisational environments. As an under- and postgraduate student I researched the effect of threat perceptions (e.g., from immigrant or religious outgroup members) on the extent to which individuals adopt more social distance (unpublished undergraduate thesis at Utrecht University) and show prejudiced behaviour (unpublished postgraduate thesis at the LSE). Based on my interest in the psychology of risk, I took up a position as research assistant at the LSE (2012-2015) to work with Dr Tom Reader on a project investigating safety culture (i.e, “the norms, values, and practices shared by groups in relation to risk and safety"; Noort et al., 2016, p.516) across the European air traffic management industry. In this highly enjoyable period, Dr Reader introduced me to a wealth of knowledge on the intersection of social psychology, organisational safety and safety management, and he mentored me in becoming a better researcher. As a side-project I worked with Dr Bradley Franks and Dr Martin Bauer on how risk perceptions (e.g., in terms of threat, uncertainty, alienation) shaped sense-making amongst conspiracy theorists. Thus, I became a published scholar on safety and risk (Franks et al., 2017; Noort et al., 2016; Reader et al., 2015), and I deeply enjoyed the mixture of writing and travelling across Europe to facilitate focus groups on safety culture.

To understand how safety culture operates in practice and gain hands-on understanding of safety management I worked at NATS (British air traffic control) in 2015-2016. Here, I was involved in projects involving safety management (e.g., safety culture, safety assurance for a new voice communication system, human performance assessment) and training (e.g., on human factors, Crew Resource Management). This period taught me two valuable lessons. First, that the world is safer with me handling theories and methods rather than planes (i.e., a hilarious simulation to familiarise myself with the work of air traffic controllers went painfully wrong). Second, that even within organisations with a strong safety culture people can withhold safety concerns. The management and operational staff I worked with (e.g., air traffic controllers, engineers, Royal Airforce Force personnel, pilots) demonstrated excellent commitment to safety, and I observed that air traffic controllers handled flights in one of Europe’s largest and most complex airspaces with skilled precision and excellent three-dimensional planning capabilities whilst using advanced technologies and following detailed protocols. In 2015-2016, these air traffic controllers handled 4,494,000 flights (i.e., aircraft taking off, landing, or passing through British airspace; NATS, 2018b, 2018a) yet only saw a single serious aviation incident that was attributable to the organisation (NATS, 2018b). However, conversations provided me with anecdotal evidence that people across the aviation industry were able to recall situations in which they were concerned but felt unable to raise this. I was puzzled by the possibility, though unverified, that even in organisations with the strongest safety records people may find it difficult to speak-up about safety. In particular, it made me curious to understand the behavioural nature of interpersonal relationships that constitute safety culture such as safety voice.

Thus, in 2016 I started with the research (undertaken for this thesis) into the behavioural nature of safety voice. This project came about after discussions between Dr Tom Reader, Dr Alex Gillespie and myself, and could be said to originate in 2012. This is when, in an LSE pub, the three of us (alongside my dear friend Adam Engstrom) first brainstormed new ways to improve safety performance through enabling better communication. In this thesis I therefore investigate the behavioural nature of safety voice because this is a culmination of my motivations, hopes, previous experiences and discussions with others. Researching safety voice is important for understanding the degree to which harm can be prevented though speaking-up about safety, and if any good may come from this work, may it be this: that someone was able to prevent harm through speaking-up because they applied insights developed in this thesis.

He alone is my refuge, my place of safety.

Psalm 91:2, NLT

ACKNOWLEDGEMENTS

The present research was made possible by the invaluable support of others. Many individuals and institutions have contributed in some form or other, and I am grateful to everyone. Of these, I would specifically like to thank:

The Economics and Social science Research Council (ESRC) and the London School of Economics and Political science (LSE) for enabling me to conduct the present research through the logistical and financial support they offered. In particular, I would like the thank the ESRC for entrusting me with a generous scholarship that enabled me to pursue this thesis.

Dr Anam Parand and Andy Stewart for providing good company over the years and advice on conducting systematic literature reviews and philosophical thought, respectively.

Lars Wicke, Dr Tamara Ansons and others at the Behavioural Research Lab for providing the research facilities and logistical support for conducting the experimental studies.

Dr Kevin Corti, Dr Brett Heasman and Dr Meghan Leaver for providing friendship, guiding me through applying to the PhD program at the LSE and illustrating excellent research practice.

The Research Ethics Committee at the LSE for carefully considering my ethics applications, enabling effective research planning through turning around applications swiftly and providing constructive feedback to further enhance ethical procedures.

Professor Dr Lucia Garcia-Lorenzo and Dr Matteo Galizzi for providing me with constructive feedback on early versions of this thesis for the extended essay and upgrade viva, respectively.

Previous scholars in the domains of social psychology and safety management: without the prior publication of your hard labour and carefully constructed thoughts I would have had less inspiration for developing the ideas in this thesis.

Editors and reviewers at the journals in which the research of thesis has been published, or to which manuscripts have been submitted. Thank you for providing the constructive feedback that substantially improved the quality of the publications.

The research assistants that have contributed to this thesis: Lindsie Arthur-Hulme, Jackie van Dael, Alex Goddard, Nanne Houtsma, Arben Islami, Sai Kalvapalle, Sidsel Larsen, Eleanor McSweeney, Jason Murcko, Celestin Okoroji, Grace Rahal, Leia Rusch, Alyssa Pandolfo, Kate Sitniewski, Rachel Taylor, Ophelia Wong, and Lucy Zhong. You proved to be worthy colleagues and provided an invaluable sounding board for initial ideas. Without you, I would still be collecting, coding and analysing data! Thank you for great company in the lab and beyond.

My viva voce examiners, Dr Ilka Gleibs and Dr Michaele Kolbe, for providing an engaging experience and challenging me on my views, arguments and conceptualisations.

My parents for supporting me throughout the years, inspiring a deep care for and knowledge about the world and for encouraging me to make the most of the opportunities provided to me.

My wife, Nelleke. For the help with the video-stimuli, providing a sounding board and because your support always carried me. Lost in study design, data collection and the write-up of this thesis, in your love and kindness you always welcomed my concerns with open arms. Thank you for being who you are and keeping me grounded.

Most of all, my gratitude goes to my supervisor Dr Tom Reader and co-supervisor Dr Alex Gillespie. I would not be where I am today without your continued mentoring and advice. My respect for you as professionals and individuals cannot be captured in words. Thank you for funding research assistants, engaging in stimulating conversations, co-authoring the presented articles and providing guidance: I have learned to have a deep appreciation for your insights and writing style!

TABLE OF CONTENTS

TABLE OF TABLES AND FIGURES

CHAPTER 1:

INTRODUCTION AND THEORETICAL BACKGROUND

The importance of safety voice for mitigating accidents

Accidents can emerge when people and institutions fail to manage hazards. Reducing the probability that hazards lead to harmful outcomes is desirable because hazards can lead to unwanted consequences such as significant financial costs (Novak, 2019), injuries (e.g., wrong-leg amputation) and fatalities (e.g., due to air crashes; NTSB, 1983). Hazards can emerge from daily life (e.g., motor accidents, trip hazards, sport injuries; Wilson, 1979), natural hazards (e.g., floods, fires, viral infections) and institutional failures in safety management (e.g., nuclear disasters, harm to patients, use of pesticides; G. W. Fischer et al., 1991; Slovic, 1987), and are considered unsafe to the extent that they provide a higher "likelihood of physical harm – whether immediate or delayed – to persons, property, or the environment" (Beus et al., 2016, p.353). A substantial literature on safety management (Beus et al., 2016; Swuste et al., 2020) has therefore aimed to understand how safer outcomes may be assured for organisations (e.g., improved safety performance) and the public (e.g., patient safety; Vincent, 2010) through the optimal design of technical, social and human factors (Swuste et al., 2020).

Research on safety management draws on a broad range of domains (e.g., management, organisational behaviour, human factors, engineering, social psychology) and contexts of application (e.g., transportation, healthcare, construction, etc.; Beus et al., 2016; Hosseinian & Torghabeh, 2012; Swuste et al., 2020; Zanko & Dawson, 2012). As Hosseinian and Torghabeh (2012) highlight, early theories such as Heinrich's Domino Theory (developed in the 1930s) proposed that accidents are caused by linear cause-effect relationships and emphasised human error (Reason, 2000). In contrast to these person-centred approaches, 'systems approaches' to safety management (Leveson, 2002; Reason, 2000) emphasise that human behaviour is part of sociotechnical systems (i.e., interactions between human, social and technical factors; e.g., Appelbaum, 1997) with human errors “seen as consequences rather than causes, [and] having their origins not so much in the perversity of human nature as in ‘upstream’ systemic factors” (Reason, 2000; p.370). For instance, this may involve the design of organisations and technical tools (e.g., ineffective training standards, protocols, tecnical specifications, etc.; Zanko & Dawson, 2012) that enable hazardous situations and unsafe acts.

Systems approaches to safety management gained prominence in the 1980s (Swuste et al., 2020) after nuclear disasters at Three Miles Island (in 1979) and Chernobyl (in 1986) prompted researchers and practitioners to better understand and manage the systemic causes of accidents. This period gave birth to seminal theories including James Reason’s (1990) Swiss Cheese model, Charles Perrow’s (2011) Normal Accident Theory, and Diane Vaughan’s (1996) normalisation of deviance. These theories stress that humans can provide a source for high reliability (e.g., through behaviourial variation enabling adaptive responses to novel threats; Reason, 2000) and argue that accidents can emerge when suboptimal behaviours are habituated (e.g., through training or observing uncorrected poor practice; Vaughan, 1996), or when typical, normal failures in the tight alignment between factors in safety management systems (Perrow, 2011) coincide in unfortunate ways (Reason, 1990).

For instance, the Human Factors Analysis and Classification System (HFCAS; Shappell & Wiegmann, 2000), a development of Reason’s Swiss Cheese model, highlights that accidents can emerge from the interplay of organisational influences (e.g., processes, climate), unsafe supervision (e.g., planned inappropriate actions, failure to correct problems), preconditions for unsafe behaviour (e.g., physical and technical issues, poor training, physical/mental limitations) and unsafe behaviour (i.e., errors and violations; Wiegmann & Shappell, 2016). This is supported by studies that indicate that accidents can emerge from common, every day behaviour of people in response to their environments (e.g., Bienefeld & Grote, 2012;

Jacobsson et al., 2012) and accident investigations that typically identify that mishaps in aviation (Tarnow, 1999, 2000), offshore oil drilling (Reader & O'Connor, 2014), space travel (Moorhead et al., 1991) and healthcare (Francis, 2013) occur when social environments (e.g., providing poor norms, values, and leadership) and technical problems (e.g., equipment failure) enable unsafe acts (e.g., the decision to proceed with the launch of the Challenger space shuttle; Vaughan, 1996).

For effective safety management, it is therefore important to consider how accidents emerge from the dynamic relationship between people's behaviour and the social and technical characteristics of hazardous scenarios. For instance, from i) the lack of hazardous scenarios eliciting risk perceptions (Slovic, 1987), ii) unfavourable organisational norms and values on coordinating about safety (Guldenmund, 2000), iii) human tendencies for deferring to authority figures (instead of contradicting them; Phelps & Reed, 2016), or iv) the ineffective flow of safety-related information when individuals are concerned about hazardous situations (Westrum, 2014). Because of this, the literature has proposed theories to explain how social environments can enable better safety performance in hazardous contexts, including safety climate (Zohar, 1980, 2010), safety culture (Guldenmund, 2000), safety leadership (Barling et al., 2002), safety participation (M. A. Griffin & Neal, 2000) and safety citizenship (Didla et al., 2009). Based on these theories, high reliability industries such as air traffic management and healthcare have established safety strategies to coordinate activities for safety management around the principle that "people create safety" (e.g., NATS, 2020; p.5; for a case study on how NATS' strategy enables the raising of concens, see Francis, 2013), or have adopted training programs (e.g., Crew Resource Management, TeamSTEPPS; Kanki et al., 2019; King et al., 2008) that enable environments in which safety-related information can flow freely (Westrum, 2014).

In particular, the concept of *safety voice* – the extent to which people can speak-up about safety concerns (Barton & Sutcliffe, 2009; Conchie et al., 2012; Tucker et al., 2008) – is considered important for safety management because it explains how accidents may be avoided through the extent to which social environments enable people to communicate concerns about encountered safety threats. This behaviour ensures the essential flow of safety-related information (Westrum, 2014), but is often challenging: risks are difficult to identify a-priori (Christensen-Szalanski & Willham, 1991), imperfect intuitions and heuristics can bias estimates of risk (Kahneman & Klein, 2009), responses to risk are socially constructed (Turner & Gray, 2009) by people with distinct access to safety information, risk appetites and levels of experience, and others may respond unfavourably (e.g., Edmondson, 1999), contest the presence of risk (Weick, 2010), or even deny that concerns have been raised (Burris et al., 2013). Crucially, and of importance for situating this thesis within the safety management literature, where safety concerns are withheld – termed *safety silence* – this can contribute to injuries, accidents and significant financial damage (Novak, 2019). This is best, and tragically, illustrated by mishaps in safety-critical industries such as space travel, oil-and-gas, aviation and healthcare that were, in part, attributable to ineffective safety voice.

For instance, in the case of the NASA Challenger space shuttle disaster (for detailed analyses, see: Moorhead et al., 1991; Vaughan, 1996), an engineer at NASA's supplier Thiokol was concerned about the cold-weather performance of a key piece of equipment (i.e., O-rings), yet was poorly listened to when he raised this concern in a safety meeting due to pressures for proceeding with the launch emerging from a contract that was up for renewal and a time-sensitive lecture from space. Whilst a concern was initially raised, the dismissal and eventual withholding of concerns about the uncertainty of the O-ring performance in cold weather contributed to eight fatalities and a temporary halt to the space shuttle program.

In the case of the blow out of BP Deepwater Horizon, a complex system of interdependent factors (e.g., poor safety culture, production pressures, inadequate regulation) led to unfavourable preconditions such as poor communication between Halliburton contractors and BP operators (Reader & O'Connor, 2014). This enabled a situation whereby safety concerns about performing a negative pressure test (i.e., indicating well integrity) were not raised. Accordingly, "those performing the [negative pressure test] did not have findings from the cement evaluation log available and were unaware of […] the need to take a more conservative approach to assessing well integrity" (Reader & O'Connor, 2014, p.422). This contributed to hydrocarbon entering the well and travelling up to the oil rig, eventually resulting in the blow out and sinking of the offshore oil rig and eleven fatalities.

In the case of the air crash of United Airlines Flight 173 (NTSB, 1978), flight crew were distracted by a loud sound and a lack of a green signal that wrongly suggested the landing gear was not appropriately locked in. However, when junior flight crew became concerned about fuel, they were unable to effectively raise this to the attention of the distracted captain because their initial concerns were dismissed (indicating poor listening to safety concerns), and safety voice only became explicit when it was too late. This contributed to a crash with ten fatalities and twenty-three persons seriously injured. Similarly, Driscoll (2002) described how poor decision-making and the lack of sharing of crucial safety information (between flight crew, or with air traffic control) contributed to the crashes of Markair Flight 3087 (resulting in three minor injuries and one severe injury), US Air Flight 1016 (resulting in three minor and sixteen severe injuries and 37 fatalities) and American Airlines Flight 965 (resulting in 159 fatalities).

Finally, the Francis inquiry into patient neglect at the NHS Mid-Staffordshire hospital trust in the United Kingdom indicated that safety silence and poor listening to safety concerns were systemic problems for the UK's National Health Service: whilst senior staff did not share this view, concerning trends were found amongst members of staff reporting they had been victimised after speaking-up about safety, or had withheld safety concerns due to a widely shared fear of being victimised. This culture, characterised by unfavourable norms and values for speaking-up, was identified as a key contributor to up to 1200 fatalities at the Mid-Staffordshire hospital trust (Francis, 2013), and led to recommendations for system-wide changes such as the installation of safety voice ambassadors that support staff in raising safety concerns (i.e., 'Freedom to Speak Up Guardians'; Francis, 2015).

Examples like these (for more cases, including on postive outcomes, see for example: Driscoll, 2002; Westrum, 2014) indicate that safety voice poses a fundamental problem for safety management (e.g., Barton & Sutcliffe, 2009): in order to overcome the momentum of hazardous situations towards harmful outcomes (e.g., through safer actions; Barton & Sutcliffe, 2009), people need to coordinate effectively on safety concerns and address distinct and imperfect risk perceptions (e.g., risky conditions, safety limits; Slovic, 1987), yet well-meaning individuals often fail to speak-up about safety[1]. Research estimates that even highly trained safety-critical staff (e.g., air crew, surgeons) only raise approximately 50% of their safety concerns (e.g., Bienefeld & Grote, 2012; Jacobsson et al., 2012). Because of this, safety management theories have tended to include safety voice within larger measurement suites (e.g., questionnaire items). For instance, i) research on safety culture has partly conceptualised dimensions for error reporting and just culture in terms of the extent to which employees feel able to raise safety concerns (Noort et al., 2016; Reader et al., 2015), ii) the 60-item Safety Attitudes Questionnaire includes nine items on teamworking that reflect the ability to

[1] This corresponds to seminal findings in the field of social psychology. As Van den Bos and colleagues (2011) pointed out aptly, the field of social psychology has long-established that because the majority of people are prosocial (e.g., people value close relationships, adhere to social norms) they do not always act on their perceptions of the situation (e.g., conformity; Asch, 1956), or behave in desirable ways (e.g., obedience to authority, the bystander effect; Darley & Latane, 1968; Milgram, 1974). In a similar way to Normal Accident Theory (Perrow, 2011), this highlights that harmful outcomes can emerge from typical, normal, social psychological processes.

effectively speak-up about safety (e.g., 'the physicians and nurses here work together as a well-coordinated team', 'in this ICU, it is difficult to speak up if I perceive a problem with patient care'; Sexton et al., 2006, p.7), and iii) studies on safety leadership (e.g., Nembhard & Edmondson, 2006) have conceptualised safety voice as a key outcome variable.

However, and despite the safety management literature recognising the importance of safety voice for preventing harmful outcomes (e.g., Barton & Sutcliffe, 2009; Conchie et al., 2012; Driscoll, 2002; Neal et al., 2000; NTSB, 1978; Reader et al., 2015; Reader & O'Connor, 2014; Sexton et al., 2006; Tucker et al., 2008), the concept of safety voice remains poorly understood. That is, whilst important insights have emerged on the themes that people use to describe their safety voice behaviour (e.g., perceived efficacy of voicing concerns, the motivation to help, hierarchies, policies and role expectations; Morrow et al., 2016) and antecedents to speaking-up (Okuyama et al., 2014), the concept remains imprecise because evidence is disintegrated and draws on distinct theoretical domains (Guldenmund, 2000; Morrison, 2011). Moreover, the relevance of safety voice for safety management remains unclear because the field i) has poorly developed how people actually raise their concerns (cf. Krenz et al., 2019), ii) has not systematically evaluated safety voice during hazardous situations posing actual risks (Krenz et al., 2020), iii) has assumed that the concept of employee voice (Morrison, 2011, 2014) can be applied to safety voice in full (e.g., Okuyama et al., 2014) or part (Tucker et al., 2008), iv) has reduced conceptual complexities by adopting positivist approaches (Kenny et al., 2020), v) has rarely established safety silence because studies tend not to assess the extent to which people are in fact concerned (cf. Schwappach & Gehring, 2014c), and vi) has provided inconsistent evidence on the success of interventions for promoting safety voice (O'Donovan & McAuliffe, 2020).

A poor understanding of the nature safety voice is problematic because without a clear and precise conceptualisation we do not know the extent to which safety voice poses a problem in seeking to avoid accidents (e.g., to what extent people actually voice concerns during hazardous situations, or across levels of analysis) and may not be able to design effective interventions for reducing safety silence (e.g., if interventions only work for specific contexts or safety concerns, this would explain mixed findings on interventions; O'Donovan & McAuliffe, 2020). Thus, in order to advance safety management theory and application, a need exists to draw together evidence on safety voice, synthesise this into a conceptual framework, clarify the behavioural nature of safety voice, explain the relevance of the concept for safety management, and identify and address gaps within the literature.

In this thesis, I therefore aim to contribute towards the field of safety management by drawing together evidence from distinct literatures that capture safety voice, and clarifying the extent to which, and how, the concept of safety voice is relevant for understanding accidents. More specifically, I i) provide a precise conceptualisation of the behavioural nature of safety voice, ii) evaluate how safety voice may be optimally assessed, and iii) appraise the implications of the nature of safety voice for designing interventions. Through four articles that emphasise conceptual and methodological aspects that relate to safety voice and safety silence, I aim to achieve this. In the remainder of this chapter, I present a broad review to scope the literature on safety voice (a systematic literature review is presented in Chapter 3) and develop three high-level research questions that capture the specific research questions developed within the presented articles (also see Table 1.1). I conclude this chapter with an overview of the thesis by delineating the intended contributions, mapping research questions onto the presented articles, and outlining the thesis structure.

Defining the safety voice concept

The concept of safety voice captures the extent to which individuals speak-up[2] about safety concerns (Barton & Sutcliffe, 2009; Conchie et al., 2012; Tucker et al., 2008) and has been explicitly defined as: "communication motivated toward changing perceived unsafe working conditions that have implications for individual and organizational health (…), [flowing] through formal and informal channels, and (…) directed toward numerous targets (e.g., supervisors/managers, coworkers, union officials, government officials)" (Tucker et al., 2008, p.320). The behaviour is intrinsically social: i) it occurs through communication that reflects people's participation in safety (Neal & M. A. Griffin, 2004), ii) is rooted in organisational and national cultural values for harm-prevention (Guldenmund, 2000; Noort et al., 2016) and iii) is reflected in group and organisational practices on safety (e.g., safety citizenship behaviours; Didla et al., 2009). For instance, individuals engage in safety voice when they communicate perceived hazards (e.g., 'I am concerned this may not be the correct patient'), safety information (e.g., 'I understood that runway was closed', 'you should not drive faster than 40 km/h through that turn'), potentially harmful outcomes (e.g., 'that stepladder might fall') and desired actions to mitigate harm ('abort!', 'look out!').

Safety voice manifests in speech (see Chapter 5) and is contrasted with the withholding of safety concerns, which, following Tucker and colleagues (2008) in adopting employee voice terminology, may be labelled 'safety silence' (compare with 'employee silence'; Morrison, 2011). Whilst research has tended to operationalise safety voice as a binary opposite to safety silence (e.g., Bienefeld & Grote, 2012; Conchie et al., 2012), the acts may be better understood as a matter of degree to which concerned people speak-up about distinct safety themes (e.g.,

---

[2] I use the hyphen in 'speaking-up' to capture the act in a single term: unlike Liu, Zhu and Yang's (2010) terminology, I do not distinguish between speaking-up' to seniors and 'speaking out' to colleagues within the terminology itself. Instead, because I acknowledge voice behaviours are target-sensitive, I specify the target of safety voice when appropriate.

safety information, stopping ongoing action; see Chapter 5). However, to date, safety silence remains poorly described because studies that capture safety voice have provided no explicit definition of the behaviour (e.g., McLaughlin et al., 2014), simply referring to ‘speaking-up about safety’ (e.g., Habyarimana & Jack, 2011), or did not explicitly distinguish the acts of raising and withholding safety concerns (e.g., Tucker, et al, 2008). For clarity, throughout this thesis, I indicate whether I refer to the concept of safety voice (i.e., capturing the broader phenomenon) or the specific acts of safety voice and safety silence[3].

Engaging in safety voice is considered beneficial for preventing harm by changing the dysfunctional momentum of hazardous situations towards harm (Barton & Sutcliffe, 2009). As discussed above and in Chapters 5 and 6, safety silence has contributed to fatal accidents across high reliability industries such as aviation, space travel, oil and gas and healthcare, with studies estimating that safety silence produces steep financial costs (Novak, 2019) and that people engage in substantially less safety voice during safety incidents than near-misses (Blanco et al., 2009). By speaking-up about safety, people can highlight risky and deficient actions (e.g., human error; Okuyama et al., 2014), establish a shared understanding on hazardous conditions (Edmondson, 1999; Reader et al., 2007), make observations actionable (Barton & Sutcliffe, 2009) and promote the identification of system weaknesses and learning from mistakes (Reader

---

[3] For instance, the ‘behavioural nature of safety voice’ refers to the concept that captures both acts of speaking-up and withholding concerns. This does not fully resolve terminological confusion because ‘safety voice’ can still refer to the act and the concept; however, I have aimed to explicitly define the behaviour throughout this thesis to address this. Arguably, terms such as ‘the extent to which concerns are raised or withheld’ or ‘the degree of safety concern communication’ might be used to accurately capture both acts, but I consider their widespread use as conceptual labels undesirable in terms of conciseness and the link with other voice literatures (e.g., employee voice). The use of ‘safety voice’ to refer to the broader phenomenon also corresponds to related fields such as ‘obedience to authority’, which capture the extent to which people obey or resist authority figures (Kaposi, 2017). Crucially, in Chapter 5, I show that safety silence is scalable in terms of the degree of safety voice for concerned individuals and this adds to the argument for conceptualising the phenomena as distinct but highly related.

et al., 2015). When others listen effectively to concerns raised this can increase safety performance (Manapragada & Bruk-Lee, 2016; Manias, 2015), prevent harm (Tucker & Turner, 2014; Turner et al., 2015), and benefit organisational outcomes (Novak, 2019; see Chapter 6). For example, safety can be maintained when junior doctors highlight to operating surgeons that a critical step (e.g., burning before cutting) has been omitted (Barzallo Salazar et al., 2014) and surgeons listen and act on this.

Theory on safety voice arose from person-centred and political studies indicating that management practices (e.g., leadership styles, commitment to safety) and team attributes (e.g., hierarchy, workload, norms for collaborating on safety) shape whether employees communicate their concerns (Hirschman, 1970; LePine & Van Dyne, 1998). The original development of the concept is attributed (Tucker et al., 2008) to the development of the Exit-Voice-Loyalty-Neglect (EVLN) framework. In his seminal work, Hirschman (1970) proposed that customers can respond to organisational decline (e.g., declining product quality) by doing business with competitors (i.e., Exit), actively raising dissatisfaction about an "objectionable state of affairs" (Hirschman, 1970, p.30) to management in charge, higher authorities or the public (i.e., Voice), or remaining silent in the face of adversity (i.e., Loyalty). Here, voice was considered a constructive step in which people actively aim to improve situations. Rusbult and colleagues (1982) and Farrell (1983) expanded this framework in terms of conceptual width and application to individuals' behaviour by proposing that romantic partners and employees can passively allow conditions to decline (i.e., Neglect).

After, approximately, 1985 (for historical reviews see: Brinsfield et al., 2009; Morrison, 2011), a second wave of voice research incorporated concepts, such as whistleblowing, in organisational behaviour into voice research (Near & Miceli, 1985). Yet, as Ma (2016) highlights, research interest in voice only became more widespread after voice was i) conceptualised as an observable and constructive extra-role behaviour (van Dyne & LePine,

1998) and ii) methodological progress was made through the development of questionnaire scales (Edmondson, 1999; LePine & Van Dyne, 1998). Whilst these two developments were arguably at odds (i.e., voice was now conceptualised as a behaviour, but new measures captured *perceptions* of behaviour), during the 2000s this led to a considerable body of research on employee voice and silence (i.e., the act of employees raising/withholding work-related issues and suggestions; Morrison, 2011, 2014) and the identification of antecedents and the conceptualisation of key processes such as psychological safety (i.e., the extent to which people expect others to respond well to voice; Edmondson, 1999), the expected utility calculus (i.e., expectations on the effectiveness of voice; Murphy & Dingwall, 2007b) and inclusive leadership (Nembhard & Edmondson, 2006).

Arguably, from the late 2000s a new wave may be distinguished whereby research has started to refine the understanding of voice by proposing sub-types (e.g., upward versus sideways, promotive versus prohibitive; Liang et al., 2012; Wu Liu et al., 2010) and facet-specific types of voice. Whilst research had been interested in cases where employees warned others about safety hazards (e.g., Barnett, 1992; Seiden et al., 2006), in the 2000s a trend emerged whereby i) safety research reconceptualised concepts from organisational psychology in terms of their nature and relevance for safety management (Didla et al., 2009; Guldenmund, 2000) and ii) voice concepts were distinguished in terms of the content raised rather than the target or motivation of voice (Brinsfield et al., 2009). Fitting this trend, the concept of safety voice emerged when Tucker and colleagues (Tucker et al., 2008) introduced the term 'employee safety voice'. As indicated above, they defined this in terms of communication with others about unsafe working conditions through diverse communication channels. In particular, employee safety voice was argued to be distinct from employee voice (Morrison, 2011) and safety participation (Neal & M. A. Griffin, 2004) because it is often protected by legal frameworks and may be understood as dissent when people challenge managers' safety accountabilities[4]. Conchie and colleagues (2012) shortened the term to 'safety voice', and this was subsequently adopted in a systematic review on safety voice within healthcare (Morrow et al., 2016).

Whilst different labels have been used to describe the behaviour (e.g., "speaking-up", "nurse voice"; Barton & Sutcliffe, 2009; Bickhoff et al., 2016; Habyarimana & Jack, 2011; Krenz et al., 2020; Manapragada & Bruk-Lee, 2016), throughout this thesis I adopt Conchie and colleagues' (2012) 'safety voice' and its counterpart 'safety silence'. This is because these terms i) are concise, ii) highlight that the nature of the behaviour involves communication (i.e., the extent to which people raise or withhold concerns), and iii) emphasise the content that is communicated (i.e., safety) rather than a target (e.g., seniors) or context of application (e.g., organisations, healthcare).

The conceptual scope of safety voice

Safety voice involves the communication of safety-related content and has a unique conceptual scope (Tucker et al., 2008). That is, safety voice should be distinguished from constructs within the domains of communication and safety, and from concepts that consider voice as collective action rather than individuals' behaviour.

First, safety voice is distinct from communication phenomena such as advocacy (e.g., Windle, Mamaril, & Fossum, 2008), whistleblowing (Near & Miceli, 1985), constructive deviance (Vadera et al., 2013), upward dissent (Kassing, 2002), patient complaints (Reader et al., 2014), and especially employee voice (Burris et al., 2013; LePine & Van Dyne, 1998; Morrison, 2014). Safety voice appears similar to these phenomena because they capture constructive communication of suggestions or dissatisfaction with perceived conditions.

---

[4] The extent to which employee voice and safety voice are conceptually related is addressed below and especially in Chapters 3 and 4.

However, the scope for safety voice is narrower than these phenomena because the concept is contingent upon people encountering safety threats and only captures communication that involves the extent to which safety concerns about hazardous conditions are raised.

Furthermore, whilst research on safety voice tends to emphasise the raising of concerns within organisation-based dyads or teams (see Chapter 3), arguably, it is not limited to organisational environments: hazards can occur across social (e.g., private transportation), natural (e.g., floods) and organisational contexts (e.g., medicine dispensation). If this proposition is accepted, safety voice has a broader context of application (i.e., incorporating non-institutional environments and stakeholders) and this distinguishes the phenomenon from concepts limited to organisational environments such as advocacy, patient complaints, and employee voice (for further comparisons with employee voice, please see Chapter 3). Moreover, raising safety concerns is more frequently protected by legal and policy frameworks (e.g., the UK employment rights act, the NHS constitution; for a review see: Mannion et al., 2018) and perceived as critical of management decisions (Tucker et al., 2008), and it has a unique set of antecedents (e.g., the speed and impact of hazards) and consequences that can be more severe (e.g., fatalities) than failures in achieving other organisational goals (e.g., maintaining financial performance).

Finally, I distinguish safety voice from post-hoc sensemaking about mistakes because safety voice has the explicit aim to prevent harm (i.e., it is future-oriented). This means that I consider phenomena such as voluntary error reporting (i.e., which have a clear aim of learning from mistakes to improve future safety) a type of safety voice to the extent that they are aimed at preventing harm instead of understanding the reasons for unfavourable outcomes. Thus, in short, in comparison to other concepts capturing communication about perceived issues, safety voice is simultaneously more specific (i.e., it only involves the communication of safety concerns to prevent harm) and broader (i.e., safety voice can occur in response to any hazard).

In contrast, safety voice should be distinguished from other safety-related concepts. Due to its emphasis on preventing harm, safety voice naturally fits with safety management research into concepts that capture the psychological processes through which regular human behaviour (e.g., conforming to social norms) can lead to human error when conditions are unfavourable (Kohn et al., 1999; Perrow, 2011; Reason, 2000). Because of this, research on safety voice draws upon a broad range of supporting safety literatures such as safety culture (Guldenmund, 2000), safety leadership (i.e., emphasising leader behaviour and influence on followers; Barling et al., 2002), safety citizenship behaviours (i.e., capturing extra-role behaviours beneficial to safety; Didla et al., 2009), and safety participation (i.e., capturing "behavior that does not directly contribute to an individual's personal safety but that does support safety in the wider organizational context"; Neal & M. A. Griffin, 2004, p.16). These phenomena capture safety-related activities (or perceptions thereof) within organisational environments and contribute important insights on the nature of safety voice (e.g., that it constitutes an important process for organisational learning on safety) and other safety-related behaviours and their antecedents (e.g., management commitment to safety; Reader et al., 2015). However, and distinguishing safety voice, these safety constructs have a broader scope because they capture phenomena beyond interpersonal communication about safety concerns (e.g., perceptions of management practices, safety behaviours such as handwashing, participation in decision-making on safety) and their effects on mitigating harm are less direct. Specifically, safety voice may warrant its own investigation besides broader safety concepts because safety voice i) directly captures the interpersonal dynamics that capture how people prevent harm through making sense of safety threats, ii) is directly contingent upon situational characteristics of discreet hazards (e.g., that elicit safety concerns, 'safety listening'; e.g., see Chapters 5 and 6) and less on trends across situations (e.g., safety leadership, safety culture), and crucially, iii) does not merely contribute to better organisational safety performance (e.g., safety citizenship) but is considered essential for mitigating accidents (as discussed above, also see Barton & Sutcliffe, 2009; Driscoll, 2002) across hazardous situations within and beyond institutional environments.

Furthermore, safety voice is relevant for classic social psychological research on bystander intervention (P. Fischer et al., 2011; Latané & Darley, 1970) and disobedience to authority (Kaposi, 2017; Milgram, 1974). These fields also aim to promote prosocial and safe behaviours in hazardous settings (e.g., harmful experimental designs), and the fields have provided findings that appear similar in terms of the inhibiting effect of authoritarian versus inclusive leadership on prosocial behaviour (e.g., Weiss et al., 2018). However, the extent to which these concepts may be applied to safety voice and vice versa remains unclear because i) few safety voice studies have framed their research in terms of these domains (cf. Bienefeld & Grote, 2012; Pian-Smith et al., 2009; Weiss et al., 2018) and ii) these concepts capture more (e.g., helping victims through intervening in fights; P. Fischer et al., 2011) and distinct behaviours (e.g., resisting continued pressure from authority figures after concerns have been initially raised), and, to the best of my awareness, have provided few insights into how people behave during actual accidents posing real safety threats.

Finally, in the context of this thesis, I treat safety voice as individuals' behaviour rather than collective action. Hirschman's original conceptualisation of voice as communication to improve "an objectionable state of affairs" (Hirschman, 1970, p.30) sparked research programs across human resource management, organisational behaviour, political science, transaction cost economics, and labour process theory (Wilkinson et al., 2014). As Wilkinson and colleagues highlighted, these domains vary in the extent to which they focus on behaviour from individuals within social contexts (e.g., human resource management, organizational behaviour) versus collective action such as union representation or communication channels (e.g., political science, transaction cost economics, labour process theory). Whilst investigating safety voice within these domains would be a valid endeavour, here I am solely interested in the behaviour of individuals engaging in interpersonal (or intragroup) communication within social settings (e.g., dyads, groups, institutions).

In summary, the concept of safety voice describes the extent to which individuals speak-up about safety concerns to prevent harmful outcomes (e.g., accidents) and draws upon distinct literatures within the domains of communication and safety. Safety voice is situated at the intersection of these literatures and this provides a distinct scope: it only involves i) communication about safety concerns and ii) safety behaviour that involves communication about safety threats. Moreover, and unlike most voice and safety concepts, the contextual scope of safety voice is not limited to organisational environments because the behaviour is contingent upon perceived risk from hazards that can occur within and beyond institutional environments.

The need to conceptualise the behavioural nature of safety voice

Because well-intended individuals can withhold safety concerns upon encountering hazards, safety management needs to design interventions for promoting safety voice/reducing safety silence[5] in order to prevent harm from mishaps (e.g., Francis, 2015; Moorhead et al., 1991; National Oil Spill Commission., 2011; Tarnow, 2000). Due to this, the investigation of safety voice has emphasised the identification of the extent to which people speak-up (Maxfield et al., 2013; Tucker et al., 2008) and the variables that promote safety voice (e.g., antecedents, interventions; O'Donovan & McAuliffe, 2020; Okuyama et al., 2014). To this end, the field

[5] For conciseness, henceforth I only refer to promoting safety voice or reducing safety silence.

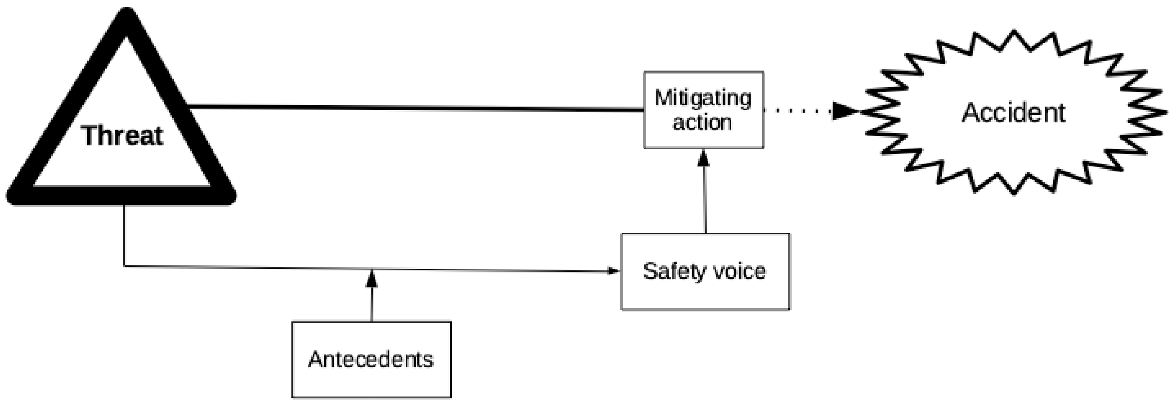


Figure 1.1. Threat Mitigation Model of safety voice upon commencing this thesis.
Note: this model highlights that the dysfunctional momentum of threats towards accidents (Barton & Sutcliff, 2009) can be mitigated through safety voice when antecedents are favourable for speaking-up (Okuyama et al., 2014). For the purpose of conciseness, antecedents are summarised.

has contributed important insights in terms of how people attribute their safety voice and silence to individual and situation-specific antecedents. For instance, to the perceived efficacy of voicing concerns, the motivation to help and institutional hierarchies, policies and role expectations (Morrow et al., 2016). Other studies have highlighted that safety voice may be promoted when people i) are aware of hazards (Lindberg et al., 2013; Manias, 2015), ii) feel a personal responsibility for contributing towards favourable situational outcomes (Aydon et al., 2016; Bickhoff et al., 2016; Jackson et al., 2010; Lyndon, 2008; Malvey et al., 2013; Manias, 2015; Nembhard, Yuan, et al., 2015; Schwappach & Gehring, 2014a), or iii) deal with leaders that act in psychologically safe ways (Bickhoff et al., 2016; Detert & Burris, 2007). Okuyama and colleagues (2014) reviewed available research in healthcare and indicated that antecedents interact with people's motive to help based on encountered safety threats. Thus, upon commencing this thesis, the conceptual model for the safety voice literature may be depicted as stating that antecedents have an impact upon the extent to which safety voice is able to mitigate the dysfunctional momentum of safety threats (Barton & Sutcliffe, 2009) towards accidents (see Figure 1.1).

However, whilst at least 48 publications have investigated safety voice (e.g., to establish antecedents, see Chapter 3), the concept remains nascent and disintegrated and we do not exactly know what the behaviour looks like in practice, or how it contributes to the mitigation of safety threats and accidents. That is, the extent to which the behavioural nature of safety voice (i.e., the nuanced manifestation of speaking-up and withholding concerns in relationship to hazardous scenarios) is important as understanding and mitigating the causes of accidents remains uncertain. For instance, it is unclear how safety voice actually occurs during hazardous scenarios beyond likelihoods of its occurrence, and whilst research on employee voice suggests that safety voice is shaped by many antecedents (Morrison, 2011, 2014), it remains unclear to what extent the content of safety concerns and context of hazards provides a different type of voice, whether safety voice is different across contexts (e.g., public versus organisational environments), or how safety concerns and safety listening (i.e., the extent to which others respond constructively or dismiss safety concerns; see Chapter 6) are important for understanding the role of safety voice in avoiding accidents. This is because studies on voice more generally (e.g., on whistleblowing; Near & Miceli, 1985) tend to adopt positivist approaches that aim to predict voice rather than conceptualise the complexities that characterise the nature of the phenomenon (Kenny et al., 2020). For instance, how its multilevel nature, relationship to safety threats, manifestation in speech, or effectiveness when others do not listen are important for conceptualising how the behaviour is relevant for preventing accidents and designing interventions. Indeed, at the start of the present research[6], evidence on the extent to which safety voice poses a problem for safety management remains limited, and it is unclear to what extent methods are appropriate for establishing the behaviour, and interventions have

[6] Where relevant literature has been published during the research for this thesis (prior to June 2020), I aimed to integrate relevant findings where appropriate. New findings were no longer included in the presented articles after these were published or submitted for publication.

had mixed success in promoting safety voice (O'Donovan & McAuliffe, 2020). In addition, conceptual and methodological assumptions remain unclear and unevaluated. For instance, whilst Tucker and colleagues (Tucker et al., 2008) conceptualised safety voice in relationship to employee voice, reviews on employee voice (Bashshur & Oc, 2014; Chou & Chang, 2017; Islam & Zyphur, 2005; Morrison, 2011, 2014; Mowbray et al., 2015) have not addressed to what extent it is appropriate to apply the construct to specific types of content (e.g., safety) or to hazards beyond organisational environments.

Thus, to better conceptualise and assess safety voice and design interventions that increase the degree and effectiveness of safety voice, research is needed to evaluate the field's constitutive assumptions and advance safety voice theory in terms of i) conceptualising the behavioural nature of the phenomenon, ii) evaluating optimal methodologies for investigating the behaviour, and iii) assessing the implications of the behavioural nature of safety voice for designing interventions. By doing this, insights may be developed into the ecological nature of safety voice (i.e., the dynamic relationship between safety voice and other variables across levels of analysis), its conceptual distinctiveness from employee voice, the extent to which the behaviour is contingent upon risk, the extent to which safety silence may be scaled upon safety voice (e.g., as a degree to which silent people engage in types of safety voice in speech), and the degree to which safety concerns (see Chapters 4 and 5) and safety listening (see Chapter 6) are essential for the effectiveness of safety voice for mitigating safety threats. Moreover, challenges for designing research and interventions for promoting safety voice (and reducing safety silence) may be identified and addressed. In this thesis I aim to contribute these insights by evaluating how safety voice behaviour should be optimally conceptualised, assessed and intervened on, and I develop these high-level and related aims in more detail below.

Aim 1: Precisely conceptualising the behavioural nature of safety voice

Research on safety voice has provided initial insights towards conceptualising safety voice behaviour. For instance, and emulating insights from research on employee voice (Morrison, 2014) and organisational safety (Guldenmund, 2000; Perrow, 2011; Reason, 2000), the field has defined and conceptualised the behaviour (Barton & Sutcliffe, 2009; Tucker et al., 2008), outlined motives for people to engage in it (Manapragada & Bruk-Lee, 2016), precisely described the likelihoods for safety voice across contexts and professional groups (e.g., Bienefeld & Grote, 2012), proposed how antecedents may elicit safety voice (e.g., Schwappach & Gehring, 2014a), and evaluated interventions for improving safety voice (e.g., Delisle et al., 2016; Habyarimana & Jack, 2011). In 2016, two systematic reviews were available on safety voice and these thematised antecedents (Morrow et al., 2016), and outlined a model for safety voice in healthcare (integrating safety voice outcomes and processes; Okuyama et al., 2014) that draws on Morrison's (2011) model for employee voice.

Yet, these systematic reviews had a limited scope (i.e., qualitative or healthcare-based research; Morrow et al., 2016; Okuyama et al., 2014). Despite drawing out important themes on safety voice and ways through which people can raise concerns in healthcare (i.e., a highly relevant domain for safety voice due to medical staff, patients and relatives frequently encountering hazardous situations), these reviews did not clarify to what degree the concept of safety voice i) may be distinguished based on its content (e.g., safety concerns), ii) may be delineated across levels of analysis (e.g., individual, group, institutional; Erez & Gati, 2004), or iii) occurs across social and organisational contexts. This is important because, as discussed above, safety voice draws upon distinct supporting literatures that use diverse labels for the phenomenon, emphasise unique contexts, use varied methodologies, may be interested in diverse antecedents and outcomes, and have distinct driving assumptions. For instance, safety voice has been distinguished from employee voice in terms of distinct contextual factors such as legal protection and safety accountabilities of leaders (Tucker et al., 2008), yet the appropriateness of applying employee voice concepts (e.g., psychological safety or the expected utility calculus; Edmondson, 1999; Murphy & Dingwall, 2007b) to safety voice remains unclear (e.g., it is unclear whether concerns about physical safety impact on interpersonal safety).

Crucially, few insights exist into the behavioural nature of the acts of safety voice and, in particular, safety silence. By adopting positivist approaches (Kenny et al., 2020), safety voice research tends to emphasise antecedent effects in favour of assessing and conceptualising how, why and whether safety voice manifests effectively. It therefore appears that the field has assumed that concepts, measures and interventions sufficiently address the extent to which well-intended individuals engage in safety silence upon encountering hazardous situations. However, this assumption may be inappropriate because we know little about safety silence. This is because studies were not designed to specifically capture safety silence, few studies have captured the extent to which silent individuals were concerned about safety, and no study has observed safety voice or safety silence directly during actual hazards (Krenz et al., 2020). Whilst it has been argued that employee silence is contingent upon people having something to say (Morrison, 2014), safety voice and silence have not been investigated in relationship to actual risk and it remains assumed that hazards provide sufficient reason for people to speak-up and that silent individuals are concerned about safety. A rare exception, Schwappach and Gehring (2014c) highlighted that people expect to speak-up more during scenarios which they rate as more harmful. Yet, they indicated the need to generalise this to behaviour by directly observing safety voice.

Thus, presently, the extent to which the behavioural nature of safety voice is contingent upon risk from hazards is unclear because we do not know the degree to which people need to be concerned in order to raise or withhold concerns, to what extent speech reflects the concerns that people have, or even if people raise concerns in response to encountered hazards. This is especially important for understanding safety silence because whilst, arguably, concerns may be gleaned from speech when people speak-up, it is unclear to what extent silent people withhold safety concerns, or if concerned individuals may express their concerns in different ways (e.g., different strengths, types of speech, across levels of analysis, over time). Addressing this is important because without evaluating how the behavioural nature of safety voice is contingent upon risk, it remains unclear to what extent research is able to capture the raising and, in particular, the withholding of safety concerns (i.e., it is unclear to what extent people withhold concerns).

Furthermore, whilst the importance of safety voice for mitigating risk is recognised and many antecedents have been described for eliciting it (e.g., psychological safety, also see Chapter 3; Edmondson, 1999), incidents keep happening despite people speaking-up (see Chapter 6). For instance, the case of United Airlines Flight 173 (NTSB, 1978) highlights that people may not effectively raise their concerns when others listen poorly to safety voice. This means that it is important to understand when safety voice is effective for mitigating harm, and whilst listening has been proposed as important for improving voice (Barlow et al., 2019; Burris et al., 2013; A. Jones & Kelly, 2014), the evidence base for this remains limited for real accidents, and a need remains for a systematic evaluation of how the concept of safety voice relates to others listening to safety voice.

Finally, it is unclear how safety voice actually manifests. Whilst safety voice can potentially manifest in distinct ways (e.g., ‘I am concerned’, ‘are you sure about doing that?’, ‘hmm, really?’), due to limited observation of the behaviour it is unclear how people raise their concerns and if there is variation in its occurrence. Whilst a few notable exceptions have directly observed safety voice (e.g., Krenz et al., 2019; Pian-Smith et al., 2009; Weiss et al., 2018), as far as I am aware, only three empirical studies have evaluated how people actually engage in safety voice (Krenz et al., 2019; Wei Liu et al., 2016; Lyndon, 2008)[7]. Yet, these established limited behavioural variation, and to my awareness no studies have addressed how safety silence manifests in speech. Conceptualising this is important for understanding to what degree safety silence is a problem for safety management (e.g., if people only withhold particular themes) and it would enable the identification of the phenomenon through alternative means (e.g., textual analysis), and targeted interventions. A need remains to evaluate to what degree safety silence may be scaled in terms of the extent to which people engage in safety voice, or as distinct types of speech (e.g., based on safety beliefs and motivations, see Chapter 5), and to propose theoretical models that capture the acts of safety voice and safety silence in relationship to risk. One way to achieve this is by directly observing safety voice and safety silence (whilst assessing safety concerns) and evaluating how thematic variations in speech exist that explain both safety voice and silence.

Thus, conceptual gaps remain in terms of the behavioural nature of safety voice. Ambiguity exists in terms of the factors that are important for understanding the behaviour (i.e., its multi-level nature, contingency upon risk, actual manifestation, adoption of employee voice concepts, extent to which others listen) and there is a need to synthesise a broader range of literature into an integrated conceptual framework. Furthermore, safety voice theory would benefit from an evaluation of assumptions in terms of the behavioural manifestation of safety voice across levels of analysis, the relationship between safety voice and safety silence, and its lack of occurrence during accidents. Addressing this is important for understanding the aspects of safety voice behaviour that need to be targeted by measures and interventions to enable individuals to raise their concerns and prevent harm. I especially address this in Chapter 3 by

---

[7] The study by Weiss and colleagues (2018) is not included here because this study investigated the effect of variation in leaders' speech (i.e., inclusiveness) on safety voice, not how safety concerns are raised or withheld.

conceptualising the ecological nature of safety voice, in Chapter 4 and 5 by identifying the importance of safety concerns, in Chapter 5 by conceptualising the manifestation of safety voice and safety silence, and in Chapter 6 by evaluating the role of actual risk and safety listening (for a precise mapping of research questions onto the presented studies, please see Table 1.1 in the section below). In doing so, I conceptualise the behavioural nature of the safety voice concept and answer the first research question for this thesis: *1) What is the behavioural nature of safety voice?*

Aim 2: Evaluating optimal methodologies to assess safety voice

Due to the emphasis on identifying variables that alter safety voice, the safety voice literature has emphasised methodologies that enable the evaluation of antecedent effects. For instance, through case studies (Seiden et al., 2006), focus groups (Malloy et al., 2009), cross-sectional surveys (Gkorezis et al., 2016), simulations (Barzallo Salazar et al., 2014; Krenz et al., 2020), vignettes (Schwappach & Gehring, 2014c) and time-lagged interventions using experience sampling (Kines et al., 2010). However, few studies have directly assessed safety voice because investigating the behavioural nature of safety voice is challenging. For instance, safety voice occurs spontaneously in natural environments and may therefore not be readily obtained, and hazardous situations are difficult to ethically introduce to participants (see Chapters 2 and 4) or compare in standardised ways (see Chapters 4 and 6). Moreover, if participants are aware of being observed this may alter their behaviour (e.g., acting in a more socially desirable manner) and the extent to which they are concerned (e.g., because they understand scenarios are simulated). Yet, because of this, few insights have been provided into the extent to which people actually engage in safety voice (or safety listening), or how the manifestation of the behaviour varies in terms of types or the strength with which safety concerns are expressed during hazardous scenarios. Additionally, studies that have directly assessed safety voice behaviour (Aubin & King, 2015; Barzallo Salazar et al., 2014; Hu et al., 2016; Hughes et al., 2014; Kolbe et al., 2012, 2014; Reime et al., 2016; Sundqvist & Carlsson, 2014) have not established the extent to which participants were concerned about safety (cf. Schwappach & Gehring, 2014c). This appears essential for distinguishing safety silence from other forms of silence (e.g., when participants have nothing to say, see Chapter 5), yet presently it is unclear to what extent studies need to assess safety concerns. Thus, because few studies have directly observed safety voice behaviour whilst assessing safety concerns, we have few insights into the manifestation of safety voice and especially safety silence.

Thus, a need exists to evaluate how safety voice and safety silence manifest, yet a gap remains because it is unclear which methodological and ethical challenges need to be addressed in order to establish safety voice behaviour through direct observation (see Chapters 2 and 4). Methodologies have specific benefits and limitations for conceptualising results (e.g., Almeida et al., 2017), with triangulation of methods compensating limitations (e.g., to address the limited generalisability of findings from experiments; Heale & Forbes, 2013), but it is not self-evident how safety voice behaviour may be optimally assessed. Challenges and limitations to obtaining data through techniques such as self-reports (Podsakoff & Organ, 1986) and simulations have been discussed (Kolbe et al., 2015; Nestel et al., 2017), but questions remain in terms of how available evidence should be optimally obtained and interpreted (e.g., what is the role of recall and imagination in self-report studies).

Evaluating the optimal way to investigate safety voice behaviour is important to enable the evaluation of the behavioural nature of safety voice and to design effective interventions. For instance, for improving training programs aimed at improving coordination on safety (e.g., Crew Resource Management training programs; Kanki et al., 2019), or revealing how the behaviour manifests for people that realise safety problems early on, but hold onto their concerns for a prolonged period (Krenz et al., 2020), or only speak-up to a small degree. When the safety voice literature relies on a single type of methodology (e.g., surveys), theoretical insights may be limited by the constraints of the methodology used and may rely on confounded insights (e.g., by a common method bias; Podsakoff & Organ, 1986). Furthermore, without clarity on optimal methods, the effectiveness of interventions may not be ascertained. Thus, a need exists for evaluating how the behavioural nature of safety voice may be optimally investigated. This might be addressed by, for instance, evaluating how experimental and field-based methodologies contribute unique insights by enabling direct access to variation in the nuanced, non-binary, manifestation of safety voice behaviour and safety concerns. Specifically, I address this in this thesis by evaluating the ethical and methodological challenges of assessing safety voice and designing an experimental methodology in Chapter 2 and 4, analysing the types and degrees of safety voice and silence in Chapter 5, and illustrating how safety voice behaviours in the field may be investigated through archival analyses of historic aviation accidents (Chapter 6). In doing so, I evaluate the second research question for this thesis, which I formulate as: *2) What is the optimal way to investigate safety voice behaviour?*

Aim 3: Evaluating how interventions should be optimally designed

Finally, the design of effective interventions may be optimally realised based upon sound concepts that theorise the behaviour targeted by the intervention. This is because interventions (e.g., to provide more encouraging leadership; Barzallo Salazar et al., 2014) may be optimal at targeting distinct subtypes of the behaviour (e.g., for concerns based on safety knowledge versus safety motivation; Christian et al., 2009) and may only be effective when certain prerequisites are met (e.g., if people are highly concerned). However, whilst studies have proposed subtypes of safety voice (Krenz et al., 2019; Manapragada & Bruk-Lee, 2016), I have not been able to identify studies that addressed the extent to which interventions are contingent upon the behavioural nature of safety voice.

Evaluating to what extent the conceptual and methodological gaps described above have implications for designing safety voice interventions is important because evidence for the success of safety voice interventions remains mixed. That is, in their review on healthcare-based interventions, O'Donovan and McAuliffe (2020) identified nine studies that used education methods combined with simulation (Pian-Smith et al., 2009), leadership videos (O'Connor et al., 2013) and case studies (e.g., H. L. Johnson & Kimsey, 2012). These studies indicated positive (e.g., H. L. Johnson & Kimsey, 2012), unfavourable (Raemer et al., 2016) and mixed effects for safety voice interventions (O'Connor et al., 2013; Pian-Smith et al., 2009). Because of this, I agree with O'Donovan and McAuliffe (2020) that further research is needed to evaluate when interventions are effective. Furthermore, the implications of the behavioural nature of safety voice for designing interventions remain unclear because the relationship between interventions and antecedents remains poorly conceptualised (e.g., in terms of mechanisms; Peadon et al., 2020) and a need remains to evaluate how safety voice is shaped by variables across levels of analysis and in terms of subtypes of the behaviour. Moreover, it remains unclear whether accidents may be better prevented by improving the likelihood of safety voice, or the extent to which others listen effectively.

Thus, in short, a need exists to evaluate the extent to which interventions are effective at improving safety voice and to conceptualise the relationship between interventions and the behavioural nature of safety voice. I address this across Chapters 3-6 and especially Chapter 5 and 6 by evaluating available evidence on safety voice antecedents and interventions and evaluating how interventions need to be tailored to the behavioural nature of safety voice. This addresses the third question for this thesis, which I formulate as: *3) To what extent do interventions for promoting safety voice and reducing safety silence need to be tailored to the behavioural nature of safety voice?*

Thesis overview

Overview of intended contributions

In summary, currently the safety voice literature emphasises that whilst safety voice is a key contributor to preventing injuries and accidents, interventions are needed because people often withhold safety concerns during safety-critical scenarios. To improve safety management (Beus et al., 2016; Swuste et al., 2020) and to design interventions to reduce safety silence, it is therefore important to consider the extent to which safety threats with a dysfunctional momentum towards accidents (Barton & Sutcliffe, 2009) elicit safety concerns, safety voice (or safety silence) and safety listening. However, currently the literature provides limited insights into how people create safety through speaking-up and a need exists to better understand safety voice in terms of how the behaviour should be optimally conceptualised (aim 1), assessed (aim 2) and intervened on (aim 3). By doing this, I intend to contribute conceptual and methodological insights specific to the safety voice literature, and to the fields of safety management and social psychology more generally. In this section, I provide an overview of intended contributions to the specific areas of research that I contribute to.

First, I intend to contribute specific conceptual and methodological insights to the safety voice literature. These include clarifying i) the relationship between employee voice and safety voice (i.e., testing a central proposition of the field; Tucker et al., 2008), ii) the relationship between safety voice and antecedents across levels of analysis (i.e., disentangling previously collapsed antecedents; Okuyama et al., 2014), iii) assumptions in theory (i.e., which remain unidentified), iv) the relationship between risk and safety voice through evaluating the extent to which people speak-up during actual hazards (i.e., because safety voice has not been studied during actual hazards; Krenz et al., 2020), v) the extent to which safety voice manifests in a distinct way (i.e., expanding the understanding of the manifestation of voice and silence in speech; e.g., Krenz et al., 2019), and vi) the importance of safety concerns and safety listening for the effectiveness of safety voice in preventing harmful outcomes. Furthermore, I aim to make tractable new ways of researching safety voice by i) evaluating optimal safety voice methodologies, ii) providing a controlled, standardised and generalisable method for investigating safety voice (i.e., an experimental paradigm), iii) providing baseline data on safety voice, iv) evaluating the need for operationalising safety voice as an observable behaviour, v) establishing the extent to which safety concerns need to be assessed to measure safety silence, and vi) illustrating how safety voice may be investigated in the field by analysing archival data.

Second, I aim to contribute to research on safety management and social psychology. By evaluating how safety voice should be optimally conceptualised, I intend to highlight the extent to which the concept of safety voice has drawn from distinct safety literatures (Didla et al., 2009; Guldenmund, 2000; Neal et al., 2000) and evaluate how safety voice may be understood as a separate behavioural phenomenon that is contingent upon situational characteristics such as risk (i.e., eliciting safety concerns) and the extent to which others listen (i.e., conceptualising an essential behaviour for the flow of safety-related information; Westrum, 2014). Thus, I aim to establish evidence that enables the proposition of a more nuanced and complete Threat Mitigation Model for the role of safety voice in the prevention of accidents (see Figure 1.1).

Moreover, by investigating the role of risk in the behavioural nature of safety voice, I aim to clarify the extent to which safety voice can shed light on how common social behaviours can inhibit safety management (i.e., testing the limits of the contribution of common social behaviour to accidents; Perrow, 2011; van den Bos, van Lange, et al., 2011). That is, I establish the extent to which people i) are willing to remain silent upon facing unfavourable leadership when their own lives are actually at stake (Kaposi, 2017; Milgram, 1974), ii) follow social norms dependent on encountered risks (i.e., expanding insights into social norms; Asch, 1956), and iii) engage in behaviours reflective of information processing during emergency situations that pose actual risk (P. Fischer et al., 2011; Latané & Darley, 1970). In addition, I contribute to the social psychological literatures on how people respond defensively to encountered threats (e.g., worldview defence in response to anxiety; Jonas et al., 2014) by highlighting that safety voice provides a specific type of defence against harm that involves sense-making for encountered hazards.

By evaluating optimal methodologies, I contribute to the safety management and social psychology literatures by indicating how behavioural methodologies (i.e., laboratory experiments, analyses of archival data originating from the field) enable access to safety behaviours that are challenging to study (e.g., in natural environments, or without exposing participants to unethical levels of risks). I also contribute a novel experimental methodology that has the potential to be adapted in order to conduct obedience research in an ethical fashion (i.e., enabling new ways of investigating phenomena such as obedience)[8].

Finally, I aim to enable more effective safety management by i) highlighting to what extent interventions have been effectively designed for people to speak-up when they are concerned about safety (i.e., explaining mixed evidence on intervention succes; O'Donovan & McAuliffe, 2020), and ii) evaluating the extent to which interventions need to address people's safety concerns (i.e., indicating the importance of risk perception for reducing safety silence; Slovic, 1987) and safety listening (i.e., evidencing the proposed importance of others' listening behaviours for effective safety voice; Barlow et al., 2019; Burris et al., 2012; A. Jones & Kelly, 2014).

[8] It should be noted that any undue pressure of authority figures should be carefully evaluated because any psychological harm to participants should prohibit the use of the experimental paradigm presented in this thesis (also see Chapter 2).

Mapping of research questions

As outlined above, in this thesis, I aim to address three aims that I summarise as: evaluating how safety voice should be optimally conceptualised, assessed and intervened on. To address this, I ask: *1) What is the behavioural nature of safety voice? 2) What is the optimal way to investigate safety voice behaviour?* And, *3) To what extent do interventions for promoting safety voice and reducing safety silence need to be tailored to the behavioural nature of safety voice?* These higher-level questions inform the specific research questions outlined and developed within the four presented articles, with multiple research questions addressing the higher level aims of this thesis (for a mapping, please see Table 1.1).

Table 1.1. Overview and mapping of research questions.

Article	**Article research questions**	**Thesis aims**		
		Aim 1: Behavioural nature	Aim 2: Optimal methods	Aim 3: Optimal interventions
Article 1 (Ch.3)	*1) Which definitions, conceptualisations and theoretical backgrounds characterise safety voice?*	X		
	2) What is the ecological nature of safety voice in terms of inhibiting (promoting) relationships between safety voice and antecedents, pragmatics, and outcomes across levels of analysis?	X		X
	3) Which methodologies have been used for researching safety voice, how suitable are they, and what methodological challenges remain?		X	
Article 2 (Ch.4)	*1) To what degree can experimental methodologies address challenges for observing the behavioural nature of safety voice?*	X	X	
	2) To what degree can experimental methodologies address challenges for reducing the methodological reliance on memory and imagination?		X	
	3) To what degree can experimental methodologies address challenges for advancing knowledge on the factors that predict safety voice?		X	X
Article 3 (Ch.5)	*1) To what degree can safety silence be scaled based on the distinct themes that people voice?*	X	X	
	2) To what degree do interventions reduce safety silence?	X		X
	3) To what degree does safety silence manifest differently over time?	X		X
Article 4 (Ch.6)	*1a) To what extent did flight crew engage in safety voice across historic aviation accidents?*	X		
	1b) To what extent has safety voice increased during aviation accidents since the 1980s?			X
	2a) To what extent does safety listening predict safety voice engagement for junior staff during aviation accidents?	X		
	2b) To what extent has safety listening improved during aviation accidents since the 1980s?			X
	3) To what degree does power distance explain safety voice and safety listening during historic aviation accidents?	X		X

Abbreviation: Ch.: Chapter.

Thesis structure

This doctoral thesis is presented as a PhD-by-publication. Four articles are presented that address the three aims of this thesis developed above through fourteen studies that use experimental and archival methodologies (see Table 1.2; methods are developed within each article, and especially Chapter 4). A narrative to connect the articles is provided through additional chapters (i.e., introduction, approach, discussion) and article prefaces. The organisation of the thesis' sections complies to guidelines outlined by LSE's PhD academy. Its general structure is consistent with past theses using the PhD-by-publication format within the Department of Psychological and Behavioural Science (e.g., using article prefaces, a scoping review alongside a systematic review). The articles are structured to meet the author guidelines for the journal in which they are published (i.e., Safety Science for Article 1, Frontiers in Psychology for Article 2), or to which they have been submitted to (i.e., Safety Science for Article 3, Risk Analysis for Article 4). For the purpose of a consistent presentation of this thesis, the complete thesis (i.e., including the published articles) has been formatted according to APA style, 7th edition and I provided a chapter number to tables and figures (e.g., 'Table 2.1' refers to Chapter 2, Table 1). I hold sole authorship for this thesis and chapters, and first authorship for the presented articles contained within it. Dr Tom Reader and Dr Alex Gillespie functioned as thesis supervisor and co-supervisor, respectively. Whilst challenging to quantify this, we agreed that as article co-authors they contributed approximately 20%. Author contributions are also specified in the preface to each article alongside the work conducted by research assistants.

Table 1.2. Overview of studies conducted as part of this thesis, in order of presentation.

Study		Date	$n_{partcipants}$	n_{other}	Format	Discussed in
1	Pilot to explore paradigms	Nov 2016	25		Laboratory	Chapter 2
2	Pilot to Walking the Plank #3 (laminate plank)	Oct 2017	8		Laboratory	Chapter 2
3	Systematic review	Mar 2017		48	Archival	Chapter 3 (main study)
4	Walking the plank #1	Apr 2017	129		Laboratory	Chapter 4 (study 1)
5	Walking the plank #2	Jun 2017	69		Laboratory	Chapter 4 (study 2)
6	Pilot to Walking the Plank #3 (info on maximum load)	Nov 2017	38		Laboratory	Chapter 4 (footnote)
7	Walking the plank #3	Jun 2018	75		Laboratory	Chapter 4 (study 3)
8	Pilot on risk perception #1	Jun 2018	88		Web-based	Chapter 4 (footnote)
9	Pilot on risk perception #2	Jun 2018	57		Web-based	Chapter 4 (footnote)
10	Pilot on risk perception #3	Jun 2018	37		Web-based	Chapter 5 (footnote)
11	Pilot to Sounds of silence study	Jul 2018	237		Web-based	Chapter 5 (pilot study)
12	Sounds of silence study	Jan 2019	404		Laboratory	Chapter 5 (main study)
13	Pilot by MSc students at LSE (closed/ open leader)	May 2017	55		Laboratory	Chapter 6 (preface)
14	Cockpit Voice Recorder study	Jan 2020		172	Archival	Chapter 6 (main study)

In the next chapters, I present my approach to conducting the present thesis (i.e., philosophy of science, how an ethical approach is achieved; Chapter 2) and four articles prefaced with a discussion regarding the relationship between the articles and the thesis' research questions. In the articles, I systematically review research on safety voice (Chapter 3), establish an experimental methodology to establish safety voice behaviours (Chapter 4), conceptualise and evaluate implications of the behavioural nature of safety voice and silence (Chapter 5) and establish the extent to which safety voice manifests during actual hazards and how this may be explained (Chapter 6). In the discussion and outlook (Chapter 7), I evaluate the contributions made to theory, methods and application, highlight limitations, and suggest new directions for research.

CHAPTER 2:
APPROACH TO INVESTIGATING SAFETY VOICE

Before presenting the studies that answer this thesis' research questions, it is important to consider how I have approached the investigation of safety voice in terms of philosophy of science and ethics. This is because, whilst a wide range of approaches to investigating safety voice are conceivable, studies can be rooted in distinct and opposing assumptions about the nature of speaking-up in response to real hazards and have the potential to inflict harm.

That is, first, because philosophy of science provides conflicting ontological (i.e., how the nature of the world is understood; Epstein, 2018) and epistemological positions (i.e., how valid knowledge about the world may be obtained; Steup & Neta, 2020), so research on safety voice may be rooted in distinct assumptions that shape how studies are designed. For instance, when the literature assumes that physical and social facts exist (Searle, 1995) and can only be known through experience (James, 1976/1910) it is more appropriate to investigate the behavioural nature of safety voice through empirical means than through thought experiments (i.e., these would not be deemed valid for providing empirical data). Conversely, if the literature assumes that the human mind is the only base of reality and the only way to know anything (i.e, idealism; Guyer & Horstmann, 2019) it is more appropriate to understand how individuals and groups think about their safety voice behaviour (e.g., through focus groups or interviews).

Moreover, due to its interests in human subjects and accident avoidance, safety voice research has the potential to inflict harm. That is, the field aims to capture the extent to which people speak-up upon encountering hazardous situations, yet when participants are exposed to, or observed during real-life hazards participants may be unduly harmed. This makes investigating safety voice challenging (i.e., it restricts the available approaches to obtaining valid data), and because of this the safety voice literature has tended to rely on self-report methodologies that ask participants to reflect on their safety voice behaviour before or after hazardous situations have occurred (e.g., Hanson, 2017; see Chapters 3 and 4) or by hiding the true intent of safety voice experiments from participants (e.g., Barzallo Salazar et al., 2014). However, empirical philosophical positions may suggest that self-reports cannot be considered optimal for drawing conclusions on behaviour (e.g., due to empirical and ontological gaps between reports and behaviour, see Chapter 4) and it is not self-evident how ethical challenges (e.g., using deception procedures versus obtaining rigorous and valid data) should be optimally addressed.

Thus, to evaluate how scholars may optimally approach the investigation of safety voice behaviour (therefore contributing to research question 2: *What is the optimal way to investigate safety voice behaviour?*), a need exists to understand how my philosophy of science provides assumptions for investigating safety voice and to evaluate how ethical challenges may be addressed by safety voice research. Without this, it is unclear how the experimental (i.e., laboratory- and web-based) and archival studies (i.e., a systematic review of historic aviation accidents) in this thesis should be designed. To address this, below I first consider the assumptions provided by my philosophy of science and its impact on designing the presented studies. Subsequently, I evaluate how safety voice research may address ethical challenges. For this, I review relevant ethical standards provided by professional bodies (e.g., the American Psychological Association, British Psychological Society; APA, 2017; BPS, 2018) and the London School of Economics and Political Science (LSE, 2016a), identify how these standards may shape safety voice research and discuss how an ethical approach is achieved in this thesis. Thus, this chapter outlines the philosophical and ethical foundation for the design of the studies in the next chapters[9]. For the purpose of conciseness, methodological challenges to investigating the behavioural nature of safety voice are not emphasised in this chapter. These are reviewed, established and addressed in Chapters 3-6 (e.g., Article 2 presents an experimental paradigm to address methodological challenges in investigating safety voice).

Philosophy of science

To reiterate, considering how my philosophy of science underpins research on safety voice is important because ontological (i.e., how I understand the nature of the physical and social world; Epstein, 2018) and epistemological positions (i.e., how valid knowledge about the world may be obtained; Steup & Neta, 2020) provide assumptions that drive theoretical and methodological choices. A full treatise on the philosophy of science in relationship to investigating safety voice, or safety management more broadly (e.g., see Haavik, 2014), is beyond the scope of the present thesis, but below I discuss how my philosophy of science impacts on the approach to investigating safety voice.

In terms of ontology, I am persuaded by perspectives that accept the existence of physical and social reality (i.e., realist ontologies), yet I acknowledge the difficulty in establishing reality as independent from human experience (i.e., idealism; Guyer & Horstmann, 2019). For instance, whilst debated for its interpretation of contexts (B. Smith & Searle, 2003), I adopt Searle's (1995) realist distinction between i) things that exist without needing interpretation such as the earth or laws of physics (i.e., 'brute facts'), and ii) things that exist within human experience such as pain (i.e., 'mental facts'). According to Searle, mental facts include a subcategory that involve things that people have agreed on to constitute a function in a context such as money, borders, or risk (i.e., 'institutional facts'). Yet, adopting pragmatic principles,

---

[9] For this, it is unavoidable that some comments look ahead to the next chapters, and I have aimed to balance the need to i) understand the choices made for designing the presented studies and ii) and maintaining the readability of this chapter for the reader who has not yet read the next chapters.

I only consider these distinctions ontologically meaningful to the extent that brute and institutional facts can elicit outcomes (e.g., hitting a brick wall, avoiding risks). As William James put it: “Everything real must be experiencable somewhere, and every kind of thing experienced must be somewhere real.” (James, 1976/1910; p.125, spelling as original). This pragmatist position does not assume that experiences (though real in themselves) need to be valid or based on something true because experiences and activities provide meaningful realities for perceivers (James, 1976/1910) and the subjective perception of environments is therefore sufficient for people to alter their behaviour.

Thus, I assume the existence of physical and social facts, but acknowledge that these are only ontologically meaningful through the consequences they can have. These consequences may be experienced differently because people are an integral component of their social, natural, and technical contexts (i.e., a holistic perspective): people dynamically shape and are affected by their environments (e.g., Erez & Gati, 2004). I therefore also accept that experiences and interpretations (e.g., the fatal outcomes of hazards and the understanding of these) can vary across contexts and do not need to be valid in order to elicit consequences. Yet, because ‘real’ resides at least in part in the experience of consequences (James, 1976/1910) I consider it problematic to deny the consequences of events in themselves (e.g., in terms of biological or social consequences such as hearts stopping, or people arranging funeral rituals)[10]. Thus, in summary, this means that for my ontological position I adopt principles from realism, pragmatism and holism, and assume that the world has a nature that is at least in

[10] To illustrate: based on our context (e.g., available information, cultural values dictating the desirability of risk and responses to risk) we may construct a hazard as risky (e.g., flying an uncertified aircraft, omitting to wash hands prior to medical procedures) and therefore alter our behaviour or proceed and experience the outcomes (none, crashing, infecting a patient). In the case of boarding an aircraft, gravity (i.e., a brute fact) and our risk assessment (i.e., an institutional fact), lead to a range of possible experiences (e.g., flying the aircraft as usual, dread, evasive actions, injuries, etc.), with the nature of the physical and social world manifesting through these experiences.

part identifiable (i.e., realism) which enables ontological claims by providing ontologically meaningful experience of consequences (i.e., pragmatism), and that the world is constituted of dynamic systems (i.e., holism).

In terms of epistemology, my assumption that reality manifests itself in the consequences it elicits means that I assume that knowledge can be advanced through empirical means. As Legg and Hookway summarise Peirce's pragmatism: "we test theories by carrying out experiments in the expectation that if the hypothesis is not true, then the experiment will fail to have some predetermined sensible effect" (Legg & Hookway, 2019; no page number). Yet, I accept that a degree of epistemological uncertainty is inherent in scientific inquiry because people are fallible (Legg & Hookway, 2019) and the process of advancing and applying knowledge is social. That is, cognitive and social biases can inhibit the advancement of knowledge, and knowledge is frequently contested because the world is represented in different ways across groups and historical periods (Bauer & Gaskell, 1999). Thus, and consistent with Haavik (2014), my epistemology is rooted in a relational social ontology: I assume that knowledge can be advanced through empirical means and acknowledge that science progresses through complex social processes of sensemaking in which scientists justify and contest knowledge (e.g., peer-reviewing).

My ontological and epistemological assumptions have at least four implications for my approach to investigating safety voice in the next chapters. It means, first, that I treat characteristics of hazardous scenarios (e.g., physical laws, safety information, technical equipment, policies) and mental states (e.g., risk perceptions) as ontologically meaningful (though distinct types of facts) because they can elicit observable consequences (e.g., speaking-up, aviation accidents). Yet, second, it highlights that I consider behaviour and perceptions different types of facts and do not assume the accurate correspondence between perceptions and the world (e.g., behaviour, safety, safety listening). Because of this, I consider self-reports meaningful (i.e., they elicit consequences) but suboptimal for establishing behaviour (i.e., an empirical gap exists between observable behaviour and mental facts). Thus, I explicitly test for the extent to which behaviour and perceptions overlap in the presented experiments (i.e., Chapters 4 and 5), and highlight the importance of assessing behaviour alongside people's perceptions about the world (i.e., Chapters 4-6). Third, my assumptions indicate that I treat empirical evidence (e.g., data from behavioural observations, reports, transcripts from conversations) as building towards a better understanding of safety voice, whilst I do not assume that evidence necessarily indicates final and optimal knowledge. This explains why in the next chapters I systematically review empirical studies (i.e., instead of opinion pieces that are valuable but present no emprical evidence; Markiewiz, 2002; Spruce, 2014), and obtain empirical data from experiments and an archival analysis (i.e., these obtain empirical data through methods that can be reproduced and refined). Finally, it means that I assume that the advancement of knowledge (e.g., raising safety concerns, building safety voice theory) is achieved through justification, contestation and refinement, which partly explains my choice to conceptualise safety listening in Chapter 6 (the other part involving its empirical and conceptual need), and to enable future research by making the study data (Chapters 3, 5 and 6), protocols (Chapter 3-5), statistical code (Chapter 5) and publications available in open access (all articles, with the same intention for Articles 3 and 4 upon publication).

An ethical approach to researching safety voice

In addition to philosophy of science, ethical challenges shape the approach to investigating safety voice. Studies have demonstrated that safety voice provides desirables outcomes because it promotes patient safety indicators (e.g., appropriate procedures; Kolbe et al., 2015), is recalled by 74-78% of professionals (i.e., trainees and 'attendings') to prevent adverse events (Belyansky et al., 2011), and may have prevented mishaps across high-reliability industries (see Chapter 1). However, whilst safety voice research has the virtuous aim of preventing harm, historic examples indicate that safety voice studies may ironically inflict harm.

In the 1960s and 1970s, research on deindividuation (as discussed in: Zimbardo, 2011) and obedience to authority, undertaken by Milgram (1963, 1974) provided valuable insights into the factors that could lead people to act malevolently. Milgram's research is particularly relevant for safety voice because of the conceptual familiarity between the conscious decisions to speak-up and disobedience when defined as an act of resistance (Kaposi, 2017). Milgram contributed a valuable conceptualisation of historical events (e.g., power relationships in Nazi Germany), yet sparked a debate on the ethics of conducting social psychological research. For instance, in terms of inflicted psychological harm (e.g., due to significant stress experienced) and the apparent lack of a right to withdraw (i.e., withdrawal was the dependent variable, and manipulations altered the perceived freedom to do so; Miller, 1986; Miller et al., 1995).

The evaluation of whether this research crossed ethical boundaries depends, in part, on guidelines available at the time (APA, 1959) and the moral philosophical position taken (Miller, 1986; for a discussion on the ethics of harm avoidance, see: Bradley, 2012; Christie et al., 2008; Miller, 1986). For example, inflicting minimal harm to gain insights into how people speak-up within malevolent scenarios may be considered universally undesirable in a deontological view (e.g., because it crosses a general principle of non-maleficence; Sharpe, 1997), acceptable in a utilitarian/consequentialist perspective (e.g., because the minimal harm weighs up against the benefit of applying new insights to prevent many accidents), or honourable within a virtue-ethics perspective (e.g., because the motive for increasing safety was appropriate for its context; Bolsin et al., 2005). Safety voice research appears to incorporate a utilitarian perspective in researching safety voice because concepts for safety voice have incorporated utilitarian principles (e.g., trade-offs between cost and benefits of speaking-up; Morrison, 2014; Schwappach & Gehring, 2014d), and the need for investigating safety voice is typically justified through its contribution to harm-avoidance (e.g., see the introduction to this thesis), which implicitly appeals to the consequences of the behaviour (e.g., Christie et al., 2008).

Yet, because of the range of perspectives on what is considered moral behaviour, and the implications of this for the approach to investigating safety voice, considering ethical standards for research design is key. Without this, participants' well-being may not be safeguarded, the field and research institutes may suffer reputation damage, and study results may be invalid (e.g., due to the altered states of participants). Thus, I now turn to reviewing ethical standards.

Review of ethical standards

Ethical guidelines have been outlined by key international and national bodies (e.g., APA, 2017; BPS, 2014, 2018), as well as the institution where this work was conducted (LSE, 2016a, 2016b). For the field of social psychology, these guidelines provide principles that serve to guide conduct and research. For example, the BPS *Code of Ethics and Conduct* (BPS, 2018) outlines professional standards to guide decision-making (1.3, 1.6)[11], and describes four ethical principles that constitute the "main domains of responsibility" (p.4) for psychologists: respect, competence, responsibility and integrity (2.1). An overlap in principles exists within and across ethics codes (Bell & Bryman, 2007), and when recent codes are mapped and synthesised, four principles provide important directions for investigating safety voice (see Table 2.1) that need to be maintained by scholars in designing and conducting research (including myself in the present thesis): respect for others, striving for excellence in research; responsibility, and minimising harm.

First, the ethical guidelines suggest that it is important to respect others. The BPS considers respect for others the "philosophical foundation for many of the other ethical Principles" (BPS,

---

[11] Numbers refer to sections of the ethics code.

Table 2.1. Principles for ethical conduct and research by ethics code.

	Respect for others	Excellence in Research	Responsibility	Minimising harm
BPS, 2018	Respect	Competence; Integrity	Responsibility	(Incorporated into Responsibility)
BPS, 2014	Respect for the autonomy, privacy and dignity of individuals and communities	Scientific integrity	Social Responsibility	Maximising benefit and minimising harm
APA, 2017	Respect for people's rights and dignity; Justice	Integrity	Fidelity and Responsibility	Beneficence and Non-maleficence
LSE, 2016	Equality of respect and Opportunity; Collegiality	Intellectual Freedom; Integrity	Responsibility and accountability	Sustainability

2018, p.5). This respect entails an understanding and appreciation of human worth and the right to self-determination, irrespective of biases that may emerge, based on "cultural, individual, and role differences" (APA, 2017, p.4), and means that no person is to be in a manner that is "less [than] favourable" (LSE, 2016a, p.3), with the benefits of the field's contributions equally accessible to all people (APA, 2017). Finally, respecting others' right to self-determination means that scholars should obtain informed consent, allow participants to withdraw consent without penalty, enable the destruction of participant data upon request, and maintain anonymity (BPS, 2018). To illustrate, respectful conduct may involve equal treatment and the inclusion of different demographics, whereas disrespectful conduct may involve harassment and discrimination against study participants.

Second, the ethical guidelines stress the importance of striving for excellence in research. Excellence in research means that safety voice scholars should strive to achieve high standards in their work, conduct themselves with integrity, and not conduct work "outside their areas of knowledge, skill, training and experience" (BPS, 2018, p.8). This includes integrity in accurately presenting research, complying with legislation (including not committing crimes), ensuring intellectual freedom and declared conflicts of interest, keeping commitments (APA, 2017; LSE, 2016a), and producing high quality research output (BPS, 2018). Furthermore, excellence in research provides for the need to obtain valid data, and thus allows safety voice studies to use deception procedures when valid data could not be collected if participants knew the study's intent beforehand (BPS, 2018). However, the ethics guidelines agree that deception should be used sparingly and only when participants are fully debriefed (i.e., explaining the study details after participation), distress from the deception is mitigated, and approval is obtained from ethics review boards (APA, 2017; BPS, 2018; LSE, 2009). To illustrate, excellence may involve the accurate reporting of a study into the influence of workload on safety voice whilst acknowledging that proposing a policy intervention to this effect includes considerations beyond one's expertise (e.g., business finance), whereas poor excellence may involve the use of deception without debriefing participants.

Third, the ethical guidelines highlight the importance of acting responsibly. Psychologists have professional autonomy due to their expertise (BPS, 2018), and this autonomy requires careful conduct around the management of entrusted data, application of knowledge and skills (BPS, 2018), management of conflicts of interest, serving the best interests of others (APA, 2017), considering research contributions to society and the potential for problematic interpretations (BPS, 2014), and raising ethical concerns when these emerge (LSE, 2016b). To illustrate, responsible conduct would be to ensure data is stored in a protected environment, whereas irresponsible conduct would be to not anonymise junior staff members' responses to a safety voice survey (i.e., to avoid identification and blame).

Finally, the ethical guidelines emphasise the importance of safety voice research minimising harm. Whilst minimising harm overlaps with taking responsibility, and especially respect for others, it is stressed by all ethics codes (APA, 2017; BPS, 2014, 2018; LSE, 2016a, 2016b). This involves minimising the impact of any variables that may negatively affect participants (e.g., stress, invasion of privacy and infringement of personal values; BPS, 2014), safeguarding participant and researchers' welfare (APA, 2017), and identifying and managing risks that exceed those encountered in daily life (BPS, 2018). Reflecting a utilitarian perspective, the

BPS (2018) considers risks that are minimal yet provide long-term benefits to be potentially appropriate, but this is for the consideration of research ethics committees. Unique amongst ethics codes, the LSE Ethics code stresses that minimising harm goes beyond minimising harm to humans and applies to the environment too (i.e., sustainability).

Maintaining ethical standards in this thesis

The four ethical principles identified within the ethical guidelines have important implications for the investigation of safety voice, and this has guided the research presented in this thesis. Because ethical challenges exist at every stage of research (i.e., from design to publication), conducting ethical research is an on-going process that has permeated all aspects of this thesis, and below I highlight how within the next chapters I have aimed to address the challenges posed by the four identified ethical principles: respect for others, excellence in research, responsibility and minimisation of harm.

Respect for others

The review of ethical guidelines indicates that, to ensure respect for others, safety voice researchers should treat participants, research assistants and the dissemination of data appropriately. To achieve this, and in addition to my personal commitment to respecting others, for the research presented in this thesis I have treated participants with dignity and in equal ways. For instance, respect for participants is achieved by recruiting across available demographics in participant pools (i.e., for the Behavioural Research Lab, Prolific Academic), without undue exclusions based on differences in culture, personality or roles (i.e., I only exclude participants if they have previously participated in any of the presented studies in this thesis). Furthermore, the ethical standards indicate that the investigation of safety voice should provide participants with the right to self-determination: they should be able to decide whether to take part and withdraw without being treated unfairly, and their data should be destroyed upon request. For the data collected in this thesis (excluding the use of publicly available archival data), this is achieved by enabling participants to make an informed decision to partake (informed consent) and requesting explicit permission to include participant's data after the full debriefing. Moreover, an additional statement is included in the debriefing information of the experimental studies: "You are completely free to withdraw your participation without any consequence, or loss of reward for your time. Please indicate this to us if this case. Your answers and video recording will then be destroyed". Future safety voice research may incorporate this statement into a debriefing procedure to ensure respect for others.

Additionally, the ethical standards indicate that research on safety voice should ensure respect for others by respecting the privacy of participants. This is achieved by collecting data in anonymous ways (i.e., names are never requested). In addition, I maintain existing levels of privacy when using archival data by not re-publishing any names of involved participants (i.e., names identified in accident investigation reports presented in Chapter 6). Finally, safety voice research should make the benefits of research equally accessible to all people, and the articles and data contained in the next chapters are therefore made publicly available (e.g., published, or submitted for publication as open access). An exception to this (i.e., public availability of the data for Article 2) is discussed below.

Respect for research assistants may, arguably, be achieved by the investigation of safety voice by treating them as colleagues in a specialised field (e.g., accepting their unique views on researching safety-related behaviours), and by providing them with a degree of self-determination in terms of the days they wanted to contribute (e.g., for collecting data). Accordingly, I communicated to research assistants that I perceived them as colleagues and provided them with flexible time-allocation. Of relevance to respecting research assistants, research that uses the Walking the plank paradigm (see Chapters 4 and 5) should address one particularly sensitive issue. Because the protocol involves a hazard whereby a research assistant walks across a plank with an alleged low maximum load, it is important to establish realism by verifying that research assistants' body weights exceed the maximum load. One's body weight can be a sensitive issue, and this indicates that safety voice research needs to identify and provide ways to appropriately address sensitive issues (i.e., I verified individual's weights but did not publish these).

In addition, safety voice research should respect research assistants by providing appropriate employment contracts. This is important for providing them protection and security, yet unfortunately, and outside my own control, delays occurred in arranging employment contracts (i.e., up to two months). This provided strains on the schedule for the thesis, and those of the research assistants. I resolved this by refusing that work would commence early, and I only accepted one occasion where research assistants volunteered to keep track of hours and submit these at a later stage. Though research assistants were duly rewarded, this provided a suboptimal solution. I addressed this subsequently by scheduling additional time for arranging contracts, and future researchers are recommended to familiarise themselves upfront with the existing contract lead times in their institutions.

Excellence in Research

Secondly, the review of ethical standards indicates that the investigation of safety voice needs to demonstrate excellence in research, and a high standard of work is achieved in this thesis by ensuring rigorous protocols and methods, accuracy in presented data, intellectual freedom and compliance with (new) legislation.

Rigorous protocols are developed in the next chapters through the iterative development of study protocols and measures. For instance, Article 2 presents three laboratory studies (supported by three pilot studies) that develop a novel experimental paradigm (labelled 'Walking the Plank') and a manual is published alongside the article (see Appendix C) that ensures a consistent execution of the protocol. Furthermore, evidence is provided on the reliability (e.g., interrater, Cronbach's Alpha) of measurement scales, application of coding schemes and developed textual dictionaries (see Chapter 5) and poor variables are dropped from analyses or subsequent studies where appropriate (e.g., for the study presented in Chapter 6, one variable with poor reliability was dropped because it could not reliably establish the pragmatic nature of messages in terms of providing information, instructing, repairing misunderstanding, etc.). Furthermore, to ensure excellence in research, the investigation of safety voice should make analyses and data verifiable. Thus, for the presented studies, I i) enable the verification of raw data by others (e.g., supervisors, reviewers), ii) provide data as supplemental material (i.e., for Chapters 3, 5 and 6), or upon request (i.e., for Chapters 4), and iii) provide a Jupyter notebook outlining statistical procedures (i.e., for Chapter 5, see appendix E).

The review of ethical standards indicates that research on safety voice should also maintain excellence in research by upholding intellectual freedom. I have therefore maintained my intellectual freedom in conducting research and writing up all sections of this thesis, and I can declare this thesis has not been unduly shaped by the institutions supporting it (i.e., London School of Economics and Political science, the Economics and Social science Research Council). Furthermore, the intellectual freedom of others is maintained through the provision of appropriate attribution (e.g., to publications, sources of the CVR data, etc.).

Excellent safety voice research adheres to available legislation, and this posed a new challenge during the time-frame of this thesis (2016-2020) because of the introduction of the General Data Protection Regulation act (GDPR; EU: 2016/679). GDPR broadly aligns with existing practices for excellent research (Chassang, 2017), and stipulates when personal data can be used for research. For instance, after explicit consent is provided for the use and sharing of data, the right to withdraw is clear, or the reuse of data is equivalent to the original purpose for collecting the data. Compliance of study protocols and data usage with GDPR is therefore achieved by obtaining informed consent, destroying data upon request and the inclusion of two explicit questions in the debriefing procedure regarding data sharing for research purposes and the permission to use video observations. GDPR came into force after the research for Chapter 4 (Article 2) was conducted, and research excellence for this chapter is therefore achieved by limiting data availability to individual researchers upon request for data verification (as is required for publications; APA, 2017).

Finally, excellent safety voice research requires that data is valid. This is challenging because informing participants about the true nature of studies can invalidate data, and deception is therefore required for the experimental studies. The permission for this procedure was obtained from the London School of Economics ethics review board (#000540), and the implications for minimising harm are discussed below. Because novel data on safety voice behaviour may necessitate deception, it is important that future researchers always engage with ethics review committees to decide how ethical standards should be balanced.

Responsibility

Third, research on safety voice should be conducted responsibly. In this thesis I achieve this by taking care in managing entrusted data, considering emerging ethical issues and highlighting study limitations and potentially problematic interpretations. For instance, collected data is stored on secure servers (i.e., with password protection, security certificates), and anonymised data is only shared freely when participants have consented. Furthermore, ethical approval was obtained from the research ethics committee at the London School of Economics and Political Science (#000540, #1051) before conducting the presented studies. Responsibility was taken by acting pro-actively upon emerging ethical concerns (e.g., applying the paradigm within an online environment, using different stimuli materials). In particular, the protocol for the laboratory-based Walking the Plank studies is designed iteratively to ensure that research assistants engage in data collection that is consistent, valid, and ethical. To illustrate, I took responsibility in developing this protocol further when it emerged that participants did not appreciate being debriefed by the same research assistant, that provided a discouraging message (a study condition). Arguably, this was because they could not sufficiently discern the behaviour of the research assistant behaviour within the study versus outside of it. After two participants responded negatively, I evaluated and altered the protocol so that participants were debriefed by a research assistant they had not previously encountered. This successfully resolved this issue as demonstrated by participants no longer raising any issues.

Lastly, in the next chapters I highlight study limitations and problematic interpretations of study results. For instance, in Chapter 6 I highlight that Article 4 should not be used to attribute blame as this requires an appreciation of systemic factors that could not be addressed within the study (also, it does not demonstrate respect for victims).

Minimising harm

Finally, safety voice studies need to minimise harm. In this thesis I achieve this by developing a protocol for the Walking the Plank paradigm that balances the imperative to minimise harm with the use of deception.

To minimise harm, the Walking the Plank paradigm presents an apparent hazard. As discussed in Chapter 4 and 6, experiments cannot investigate safety voice through the introduction of actual hazards as this practice exposes participants to inappropriate risk. For example, during a brainstorm, I briefly considered the idea of investigating safety voice in a scenario where participants could speak-up about a researcher mindlessly crossing a busy road on campus but abandoned this because the extent of risk could not be controlled. Similarly, I discontinued a pilot study after only eight participants because it became clear that I could not guarantee minimal harm. That is, to better understand the limits of perceived risk on safety voice, in an early pilot of the Walking the Plank paradigm participants were told to test the idea of a bench by sitting down on a highly breakable (i.e., laminate) plank. Participants were not exposed to actual risk (i.e., as required for the ethics approval, I intervened before they sat down), but seven of the eight participants moved to sit down with surprising speed (and I did not trust myself to always be prepared). To overcome the challenge posed by the need to minimise harm, safety voice studies may therefore better present apparent hazards that elicit the perception of risk. However, and crucially, for safety voice research, understanding the true nature of studies (e.g., that they are safe and intended to observe voice) would eliminate risk perceptions and reduce research excellence (i.e., valid behavioural data on safety voice and silence requires that participants perceive risk).

This means that because risk perceptions are necessary for observing safety voice and silence in laboratory settings, I use deception procedures (see Chapter 4). The use of deception provides a challenge to psychological research because it reduces participants' autonomy to decide whether to participate, and can cause minor psychological harms (e.g., feeling tricked). For the Walking the Plank studies, participants are recruited to a so-called 'creativity study' and the use of this deception received ethics approval because of the absence of actual risks, the long-term benefits of the research (e.g., design of safety interventions), the provision of a full debriefing and close monitoring for ethical concerns being raised. For instance, the protocol specified that research assistants should immediately follow-up with me when ethical concerns were raised. In two cases where this happened for an early pilot, this led to the protocol being changed so that the debriefing was conducted by another research assistant (as discussed above). As Corti (2015) highlighted, whether deception procedures are appropriate for experimental studies can, in part, be inferred from participants' response to the deception. Accordingly, to ensure minimal harm is achieved, I included one question for the debriefing of the study presented in Chapter 5 (n = 404): 'would you allow future participants to take part?'. Of the four participants (i.e., 1% of the respondents) that said they would not allow future participants to take part in the study, three indicated they had provided the wrong answer (e.g., "Oops. I think I pressed the wrong option. I would allow future participants to [take part], it's a really interesting study") and the final participant suggested that the trade-off with gaining knowledge would justify the deception (i.e., "if it's towards research and understanding human behaviour and how it truly is then why not"). This indicates harm was successfully minimised and illustrates how future research may evidence this.

Furthermore, minimal psychological harm for research assistants was achieved by providing safeguarding to participants coding the data for Chapter 6. This data involves the communication of people involved in fatal scenarios, providing the potential for undue stress by being deeply involved in the data. Indeed, when pilot coding the data, I felt it was a saddening experience and ensured frequent brakes and discussions with others. Thus, and on the suggestion of the research ethics committee (personal communication, 27 January 2020), research assistants' wellbeing was safeguarded by discussing the impact of data coding at regular meetings, encouraging the discussion of their experience with others, and allowing completion of data coding at their own pace (i.e., to reduce time-pressure).

Finally, I achieved the minimisation of harm by only providing paper-based study information (for participants to take home) upon request, obtaining a reusable mug for coffee breaks, and moving towards electronic data collection.

Additional ethical challenges

Investigating safety voice is complex due to its relationship to hazards and dependence on human participants, and the phenomenon can be investigated through various methods (e.g., simulations, surveys, participant observation, archival data) and in different environments. This means that, in addition to the challenges associated with maintaining the ethical standards in experimental and archival studies (as discussed above and in Chapters 3, 4 and 6), further challenges can be identified that are relevant for this thesis and should be addressed by future research on safety voice. These ethical challenges are associated with online data collection, in-situ observations, and obtaining data within organisational environments. Because these challenges are relevant for the optimal approach to investigating safety voice (research question 2), I discuss these below. Yet, I do not provide an exhaustive account because this is beyond the scope of this thesis and ethical challenges to safety management research have been discussed elsewhere (e.g., Whicher et al., 2015).

Online studies: fair pay and exit options

Because safety voice is difficult to observe in-situ, and simulations and laboratory experiments are resource intensive, online methods have been developed to enable resource efficient data collection (e.g., online surveys, vignettes). Yet, to adhere to the ethical guidelines, online methodologies should encourage ways for participants to make autonomous decisions, exert control over study withdrawal, and obtain fair reward. That is, in order to respect participants during online studies, scholars should provide autonomy when making the decision to participate (i.e., informed consent), provide the right to withdraw (i.e., providing autonomy and control) and should avoid asymmetrical power relations (e.g., through one-sided rejection options; Gleibs, 2014, 2017).

These challenges are relevant to this thesis, as pilot studies collected data through Prolific.ac: an online platform for participant recruitment (Palan & Schitter, 2018). Palan and Schitter highlighted that Prolific.ac can address challenges in terms of providing autonomy, control and fairness. That is, ensuring autonomy, Prolific.ac requires that participants receive informed consent, and this was achieved by explaining that participants would evaluate a video from an experiment. Ensuring control, Prolific.ac provides participants with the option to withdraw consent for the study without affecting their acceptance score (i.e., a reputation score) by terminating the study, letting the study time-out, or appealing a rejected submission. Furthermore, the presented studies enabled participants to skip questions (i.e., providing additional control and autonomy) by prompting for, but not requiring answers. Finally, ensuring fairness, Prolific.ac requires a minimum reward per hour (i.e., £5) that participants are made aware off when signing up to the platform (Palan & Schitter, 2018), and thus participants are fairly rewarded. The ethical challenges associated with online data collection are therefore addressed within the design of the online platform, and future researchers should carefully consider whether web-based data collection methods enable autonomy, control and fairness (e.g., for a discussion of ethics in online environments, see: Gleibs, 2017, 2014).

In-situ observations: observing versus intervening

Safety voice occurs spontaneously in response to encountered hazards. Yet, obtaining valid data on safety voice behaviour in-situ (i.e., real-time, as hazards emerge) is challenging. That is, researchers could be exposed to risks exceeding those of daily life when observing safety voice in-situ, because even highly trained observers with good subject-matter expertise may have a different situational awareness than the observed actors. Furthermore, to minimise harm, scholars cannot initiate unsafe actions or prolong hazardous situations. For example, when hazards emerge (e.g., low aircraft altitude; absence of handwashing), observers may notice that actors do not respond. This prompts a need to decide on an ethical dilemma for minimising harm that resembles the trolley dilemma in its choice of deontological versus consequentialist ethics (Bruers & Braeckman, 2014). Scholars could either i) intervene to mitigate harm, yet prohibit knowledge relevant for future safety voice interventions, or ii) observe the behaviour whilst allowing for potential harm. This is a catch-22 as both choices can prevent harm (i.e., immediate vs long-term), and providing an answer based on moral intuition is not straightforward due to the potential for cognitive biases (Bruers & Braeckman, 2014). Despite this, and arguably, intervening in this scenario may be desirable due to the uncertainty in terms of the long-term harm that could be averted. Non-intervention may be unethical because the long-term benefits of non-interventions are difficult to predict and can be substantially costlier than arranging alternative data-collection (e.g., physical harm is costlier than arranging alternative observation opportunities). The ethical guidelines suggest participants may be exposed to risks that do not exceed those of daily life, and the benefits of understanding safety voice may be substantial (e.g., by avoiding future mishaps). Yet, whilst this argument on psychological uncertainty has flaws (e.g., uncertainty also exists for whether interventions avert harm; Bruers & Braeckman, 2014), if it is accepted, it means that safety voice cannot be readily observed by observers present in-situ as hazards would eventually require that observers intervene, thus invalidating data on safety silence (i.e., would someone have spoken-up after the observers' intervention?). In this thesis, I addressed this challenge by obtaining in-situ data from historic situations (i.e., Cockpit Voice Recorder transcript in Chapter 6) and future research may address this ethical challenge by exploring how the moral philosophy underpinning related concepts (e.g., virtue-ethics and whistleblowing; Bolsin et al., 2005) may apply to designing safety voice studies.

Organisational environments: anonymity, consent and equal benefit

Whilst hazards occur across contexts, Chapter 3 highlights that safety voice research has mainly used data from organisational environments. Researching safety voice within these contexts needs to maintain the same ethical standards, and unique challenges exist in terms of ensuring anonymity, requesting consent and providing equal benefit to study participants.

The study of safety voice in organisations requires that data is collected on whether employees speak-up, for example through employee self-reports or supervisor statements. The ethical guidelines suggest that to respect the dignity of others, study participants should be granted privacy in terms of the data they provide (e.g., BPS, 2018). This is important for investigating safety voice because participants demonstrate or disclose behaviour that may be considered undesirable (e.g., safety silence), and this can lead to unintended consequences. To minimise harm, and whilst favourable outcomes are also possible (e.g., promotion), researchers should therefore anonymise participants to ensure minimal harm (e.g., demotions, ostracism). Yet, achieving this may be challenging for organisational environments because even deidentified data may reveal individual employees (e.g., due to their unique role, or known views), and it is unclear to what degree the public role of seniors warrants privacy, or whether it is justifiable to reveal organisations included in datasets (Bell & Bryman, 2007). This underscores the importance of future research on safety voice carefully considering the potential consequences for employees of disclosing their information, and the ethical guidelines suggest scholars may mitigate this by providing leaders and organisations with the autonomy to take informed decisions on whether they want to be identified.

In addition, future research should evaluate whether sufficient autonomy is provided when consent is obtained through gatekeepers (e.g., managers approving employee surveys). Gatekeepers often have the authority to make organisational decisions, yet it is unclear whether these can provide consent for individual employees, or if this leverages undue organisational pressure on individuals to participate (BPS, 2018). Accordingly, when consent is obtained through gatekeepers, future research should carefully consider whether it may be appropriate to provide additional information and autonomy to employees to ensure informed consent.

Finally, it is important to consider whether organisational stakeholders (employees, management, patients, etc.) receive equal benefit from safety voice research. For instance, when safety voice interventions include control groups (Kines et al., 2010) participants may not receive equal benefits (e.g., protection from harm). Future safety voice interventions within organisational environments may therefore utilise stepped wedge trial designs (i.e., the phased introduction of interventions for all participants; e.g., C. A. Brown & Lilford, 2006).

Conclusions on my approach in this thesis

The investigation of safety voice aims to improve safety management by enabling people to speak-up about safety, and to achieve this research needs to obtain valid data and comply with ethical guidelines. In this chapter, I reviewed the implications of my philosophy of science and the ethical challenges in designing research on safety voice, and especially the studies in the next chapters. By doing this, I revealed that the approach in this thesis is shaped by at least four assumptions that emerge from a mixture of realist, pragmatist and holistic philosophies, namely that i) characteristics of hazardous scenarios and mental states are meaningful through their observable consequences, ii) mental states do not necessarily correspond to the physical world or behaviour, iii) empirical evidence contributes towards a better understanding of safety voice, that is not final, and iv) research on safety voice is a product of on-going justification, contestation and refinement.

Furthermore, I identified four principles for conducting ethical research (i.e., respect for others, excellence in research, responsibility, minimising harm), evaluated the present thesis in terms of these principles and emerging challenges, and identified additional challenges for future research on safety voice. The evaluation of ethics for safety voice research is important, as demonstrated by the ethical controversies in the 1960s and 70s, and the ethical guidelines demonstrate that the field of social psychology has learned important lessons that guide contemporary researchers (e.g., the right to withdraw without punishment). By reviewing the ethical standards, I was able to identify the challenges involved in investigating safety voice. Together, this chapter underscores that, despite its aim of contributing valid knowledge to the field of safety management, research on safety voice is shaped by philosophical assumptions and can inflict harm: scholars should continually monitor whether the process of contributing knowledge on safety voice achieves justifiable knowledge and ethical standards. The next chapters present studies that illustrate how the field of safety voice may achieve this.

CHAPTER 3:

SYSTEMATIC REVIEW OF THE SAFETY VOICE LITERATURE

Preface

In the first article I conducted a systematic literature review to establish the gap in the safety management literature in terms of the need to advance safety voice concepts, assessments and interventions. The decision to systematically review research on safety voice was taken because upon taking stock of the literature it appeared that research on safety voice remained disintegrated, and it was unclear whether safety voice had a coherent conceptualisation and operationalisation. Studies adopted different labels (e.g., “voice”, “safety voice”, “speaking-up”, “nurse voice”; Krenz et al., 2020), definitions and operationalisations for the act of speaking-up about safety, and have drawn on distinct domains, including employee voice (e.g., Morrison, 2011) and safety culture (Guldenmund, 2000). Moreover, whilst studies had applied employee voice concepts to safety voice (e.g., Schwappach & Gehring, 2014d), it was proposed that the phenomenon is distinct (Tucker et al., 2008) and it was unclear to what extent this was appropriate. Thus, a need existed to draw together evidence from across safety management and voice research and evaluate the degree to which safety voice i) is conceptually related to the concept of employee voice (e.g., is safety voice a subtype or merely highly related to employee voice?), ii) occurs across levels of analysis, and iii) is optimally investigated. By doing this, a better understanding may be enabled of how the behavioural nature of safety voice is important for understanding accidents, and how gaps within the literature may be identified.

In Article 1, I therefore aimed to understand the state of the literature and identify gaps that I address in the subsequent studies. The article builds towards answering the thesis research questions by clarifying how safety voice is defined, conceptualised and delineated across levels of analysis (therefore contributing to research question 1: *What is the behavioural nature of* *safety voice?*), identifying methodical practices and remaining challenges (therefore contributing to research question 2: *What is the optimal way to investigate safety voice behaviour?*) and identifying and integrating evidence on the antecedents that have been established in relationship to safety voice (therefore contributing to research question 3: *To what extent do interventions for promoting safety voice and reducing safety silence need to be tailored to the behavioural nature of safety voice?*).

Study description

The article presents a systematic literature review in which I use an explicit search strategy and criteria for including or excluding studies (copies of these are provided in Appendix A). The protocol for the search strategy was developed iteratively (the versions of the search history are presented in Appendix A) to ensure coverage and a manageable dataset. To ensure coverage across theoretical domains, I decided to use two search engines (i.e., pubMED and EBSCO host) that enabled a search across diverse publication databases in social science, healthcare, legal science, history, and business. Initial exploratory searches revealed that results included irrelevant studies on consistent themes such as voice technology, the anatomy of the vocal cord, therapy for people hearing voices, and studies that included people from diverse languages (e.g., 'English-speaking'). These terms were therefore used to build the exclusion criteria. Furthermore, this aided in reducing the unwieldy number of hits (i.e., 20,655). The first formal search was still limited to organisational environments: I decided to drop this search criterion as it did not enable an evaluation of the extent to which safety voice was investigated across social and organisational environments. The final search provided 3,031 peer-reviewed publications in English. By screening titles and abstracts and accessing full-text publications, these were narrowed down to 48 articles (with 50 studies). I extracted and synthesised data from these studies to understand the behavioural nature of safety voice, compared the concept of safety voice with employee voice, and evaluated the extent to which safety voice has been investigated using varied methodologies.

A secondary aim, though not explicitly described in the article, was to contribute towards open science. To achieve this, I pre-registered the search strategy on PROSPERO (Noort et al., 2017), added open science to the assessment of publication quality, and provided the study data as supplementary material to the publication of the article in Safety Science (Noort et al., 2019b). Thus, I provided clarity on the development of the protocol and an advance notice to the field. This proved useful to avoid unnecessary duplication in the scope of reviews in the wider literature: during the process of writing the systematic review I was contacted by Dr Aled Jones to discuss how he might avoid duplication in a new review (personal communication, 31 May 2018).

In the article's discussion, I highlighted that the systematic literature contributed insights for conceptualising and operationalising safety voice. These include the insight that the concept of safety voice i) has unique scope in comparison to employee voice through its emphasis on harm prevention (i.e., underscoring its importance for the safety management literature), ii) is

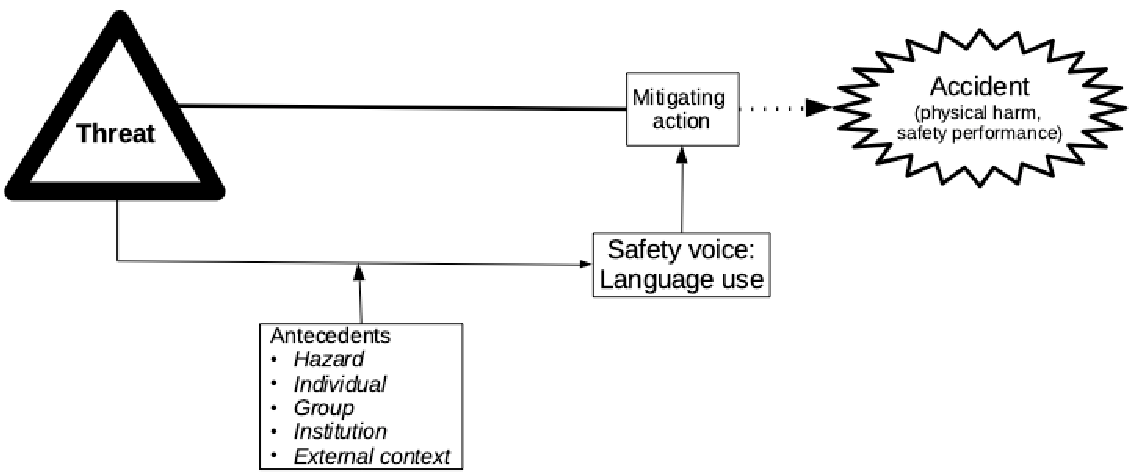


Figure 3.1. Threat Mitigation Model of safety voice.
Note: this model highlights that the dysfunctional momentum of threats towards accidents (Barton & Sutcliff, 2009) can be mitigated through safety voice that has antecedents and outcomes across levels of analysis.

ecological in terms of occurring across levels of analysis, iii) is assumed to be intrinsically social, beneficial and organisational, and iv) relies on evidence from post-hoc and report-based methodologies (e.g., interviews, surveys) that are assumed to reflect the behaviour. I proposed an ecological framework that conceptualises the behavioural nature of safety voice as situated across levels of analysis, and this advances the conceptual model for safety voice (see Figure 1.1) by indicating that the level of analysis can reveal distinct antecedents, voice and outcomes (see Figure 3.1). Finally, I identified new directions for the safety voice literature.

The ecological nature of safety voice is important for safety management because it highlights that safety voice may be conceptualised in holistic terms (i.e., emphasising that behaviour is situated in context and history; Fay, 1996), may be mapped unto multilevel models (e.g., Reason, 1990; Shappell & Wiegmann, 2000) and provide insights into interpersonal communication about safety within sociotechnical systems (e.g., Appelbaum, 1997; Leveson, 2002; Reason, 2000).

Authorship

I was responsible for designing and conceptualising the study, data collection, formal data analysis and interpretation, and manuscript preparation (80%). Dr Reader and Dr Gillespie contributed to refining the study design, interpreting results, and reviewing and editing the manuscript (20%). Anonymous reviewers provided invaluable challenges and suggestions for improving the manuscript, and Jackie van Dael contributed to the intercoder reliability analysis.

Article 1:
Speaking-up to prevent harm – A systematic review of the safety voice literature

Published under a Creative Commons licence (CC-BY 4.0) in Safety Science as:

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*, 375-387. https://doi.org/10.1016/j.ssci.2019.04.039.

Abstract

Safety voice is the act of speaking-up about safety in order to prevent accidents and physical harm. It occurs across contexts (e.g., healthcare, aviation, construction, mountaineering, high-risk sports) and understanding the phenomenon enables interventions. Despite recent interest, however, it remains unclear how safety voice i) differs conceptually from employee voice, ii) is delineated across levels of analysis, and iii) could be optimally investigated. Addressing this, we identified 48 articles, and integrated 256 safety voice antecedents, 7 pragmatics and 23 outcomes into an ecological framework. Overlap was found with employee voice concepts and methodologies, especially for the behavioural nature of speaking-up. Nonetheless, safety voice appeared unique in terms of the content of the raised message (e.g., limited to safety), the context and person speaking-up, identified antecedents (e.g., hazard-specific antecedents), and methodological challenges (e.g., operationalisation of victimhood). Our proposed safety voice framework provides a novel approach to safety voice that is ecological and indicates interventions for mitigating physical harm.

Keywords: Safety voice, Speaking-up, Employee Voice, Systematic review, Harm prevention

Article highlights:

- Safety voice is the act of speaking-up about safety to prevent physical harm.
- Conceptually related to employee voice, safety voice has unique scope due to its emphasis on physical harm.
- It is ecological: predictors, pragmatics and outcomes are found across levels of analysis.
- Important challenges for safety voice remain in developing methodologies and interventions.

Introduction

Safety voice is the act of speaking-up to prevent physical harm from hazardous situations (Bienefeld & Grote, 2012). Hazardous situations permeate organisations (e.g., dispensing medication, operating heavy goods equipment) and daily life (e.g., driving, high-risk sports; G. W. Fischer et al., 1991; Wilson, 1979), and raising safety concerns can identify and prevent potentially disastrous outcomes from these (e.g., medication error, crashes, drowning). Popular discourse frequently attributes the causes of mishaps to a lack of safety voice (BBC, 2015), and safety voice is repeatedly shown as an antecedent to avoiding harm (Turner et al., 2015b).

The role of safety voice in accident prevention has led to considerable research interest, with observations, surveys, and interviews being used to investigate the antecedents, pragmatics, and consequences of raising safety concerns in various domains, and organisational environments in particular (Morrow et al., 2016; Okuyama et al., 2014). However, it remains unclear (i) how or why safety voice is conceptually distinct from phenomena such as employee voice, (ii) how levels of analysis (e.g., individual, team, organisation) at which safety voice operates are delineated, or (iii) what the optimal methodological approach to studying safety voice is. We conduct a systematic review to assess the uniqueness of the safety voice concept, integrate the safety voice literature into a conceptual ecological framework (i.e., a model outlining antecedents, pragmatics and outcomes across levels of analysis), and consider the methodological approaches best suited to studying safety voice. Through exploring safety voice concepts and methodologies, we provide clarity on the conceptual nature of safety voice, its ecological nature and methodological challenges for research, and outline possible future directions.

Safety voice: The need for an improved conceptualisation

The concept of safety voice describes acts of communication aimed at preventing physical harm through communicating safety concerns to others. Various definitions of safety have been proposed. For example, Tucker and colleagues (2008) stated that safety voice: "(a) is communication motivated toward changing perceived unsafe working conditions that have implications for individual and organizational health, (b) can flow through formal and informal channels, and (c) can be directed toward numerous targets (e.g., supervisors/managers, coworkers, union officials, government officials)" (p. 320). Other characteristics of safety voice have been considered: for example, its constructive (Hu et al., 2015) and challenging nature (H. L. Johnson & Kimsey, 2012), or its occurrence in improving general safety (A. Jones et al., 2016) versus emergency situations (Schwappach & Gehring, 2014d).

The interest in safety voice has generated considerable research, with two reviews addressing the antecedents and consequences of safety voice (Morrow et al., 2016; Okuyama et al., 2014). Conceptually, and in particular for healthcare workers, these reviews highlighted that (i) employees report a hesitancy for raising safety concerns, (ii) predictors are contextual (e.g., a scenario causing a motivation to speak-up, leadership, work relations) and individual (e.g., felt responsibility, a cost-benefit analyses of effectiveness and psychological safety), and (iii) raising safety concerns can avoid physical harm (e.g., through error correction). Yet, the safety voice concept remains disintegrated and nascent, with three outstanding issues.

Clarifying the conceptual relationship between safety voice and employee voice

Research on safety voice draws from work in organisational behaviour on employee voice/silence, yet is distinguished through its focus on safety (Tucker, Chmiel, Turner, Sandy Hershcovis, et al., 2008). The concept of employee voice is used to study discretionary suggestions by employees that are intended to improve work-related issues (Morrison, 2014) such as smoother procedures, innovations or halt of organisational decline. The origins of employee voice are attributed to Hirschman (1970) who investigated how organisational stakeholders dissatisfied with organisational decline chose to exit a company, stay loyal or voice their concerns. Subsequent research conceptualised employee voice as the neglect of issues (Farrell, 1983) and a time-bound, observable, extra-role behaviour (Van Dyne & LePine, 1998), and used questionnaire scales to measure it (LePine & Van Dyne, 1998). Thus, research attempted to identify the antecedents to employee voice, conceptualise its processes (Milliken et al., 2003) and identify sub-types of voice (Liang et al., 2012; Wu Liu et al., 2010).

Employee voice and safety voice overlap conceptually because both refer to extra-role communicative acts to address perceived issues and change the status quo (Manapragada & Bruk-Lee, 2016; Morrison, 2014). Thus, it is important to integrate these conceptualisations, whilst addressing the ways in which safety voice is distinct (Wilkinson et al., 2019).

First, the phenomenon of safety voice is broader in sampling (i.e., it goes beyond employees, for example to patients in hospitals, or passersby) and narrower in phenomenology (i.e., focusing on preventing harm). Unlike employee voice, raising safety concerns can be a legally required and protected activity (whistleblowing; Tucker, Chmiel, Turner, Sandy Hershcovis, et al., 2008). Yet, due to the social risks involved, engagement in safety voice can be highly challenging. For example, research in healthcare has long examined why clinicians are hesitant to report observing concerns about safety, and these have included cultures of blame (e.g., Waring, 2005), non-receptive colleagues (A. Jones et al., 2016), fear of negative repercussions

(Manapragada & Bruk-Lee, 2016), and, unlike employee voice, aspects of the incident (e.g., speed of the incident, Schwappach & Gehring, 2014d).

Second, the consequences of an absence of safety voice can be severe and highly proximal, with consequences for self (e.g., personal harm) and others (e.g., organisational accidents). For example, in a decision-making analysis of the Challenger space shuttle disaster (Moorhead et al., 1991), self-censorship of supplier staff members was deemed an important factor leading to the eventual seven fatalities (i.e., after pressure from NASA, an earlier held safety concern was suppressed in the final recommendation for launch). Similarly, analysing communication in healthcare, Wei Liu, et al. (2016) showed that raising safety concerns mitigated medication errors. Research on employee voice tends to focus on events with more individualised, and less severe, outcomes that carry a lesser moral obligation for raising issues. For example, a lack of employee voice can lead to unfavourable outcomes in terms of job satisfaction, turnover, citizenship behaviours or organisational performance (Bashshur & Oc, 2014).

Third, safety voice research is grounded within a distinct set of literatures to employee voice. For example, the safety culture and climate literatures (e.g., Gauld & Horsburgh, 2014), and research on human error and systems theory (Aydon et al., 2016; Barton & Sutcliffe, 2009). These fields describe how characteristics of social systems (or lack thereof) enable safety, with speaking-up on safety frequently being incorporated into measures of safety culture and climate (Reader et al., 2015; Sexton et al., 2006) and safety citizenship (e.g., "I make suggestions to management to improve the safety of the work environment"; Reader et al., 2016, p.9). Furthermore, due to its social nature, safety voice has been framed and investigated through research on interdisciplinary collaboration (e.g., P. Liu & Ma, 2016), crew resource management (e.g., Lyndon, 2008) and shared decision-making (e.g., Frosch et al., 2012). This has helped identify group and institutional antecedents to safety voice (e.g., different experience levels; Wei Liu et al., 2016).

Fourth, within safety-critical environments, extra-role behaviours can be empirically distinguished on whether they are safety- or organisation-specific. Voicing safety concerns is considered a safety citizenship behaviour (i.e., extra-role behaviours for managing risks; Didla et al., 2009), and this wider concept is shown to be distinct from organisational citizenship behaviours (Reader et al., 2016). Direct empirical tests comparing safety and employee voice remain absent, and are beyond the scope of this article, yet this suggests that within safety critical contexts, the safety-related content of the communicated message may provide unique practical relevance beyond employee voice and a need to investigate the content of the message voiced (Morrison, 2011; Wilkinson et al., 2019).

Thus, in terms of conceptualising safety voice, a key observation emerges. Initially, safety voice appears to be a similar phenomenon to employee voice. It involves discretionary acts of communicating issues, to those with institutional power, in order to improve the status-quo. However, it also appears distinct, with relevance to those outside of an organisation (e.g., the public), different triggers to voice across levels of analysis (e.g., legal necessity, personality, observing hazards), unique consequences (e.g., personal harm, accidents), practical relevance within safety-critical organisations, and distinctive foundational literature. Therefore, it is not clear whether safety voice should be considered a subtype of employee voice phenomena, or a unique concept drawing on overlapping ideas and behaviours. This is important, but has not been directly addressed in reviews of the concept, with models of safety voice using models of employee voice to thematise research findings (Morrow et al., 2016). Yet, if safety voice is a sub-type of employee voice (Morrison, 2011), the need for a distinct literature is lessened with considerable scope for integration (Wilkinson et al., 2019), and the research findings (e.g., on voice antecedents) and methodologies for studying employee voice can be assumed to apply to safety voice. Alternatively, if safety voice is indeed an overlapping but highly distinct phenomenon, then the distinctive scope of this domain (e.g., outcomes), key research findings (e.g., different relationships amongst safety voice antecedents, pragmatics and outcomes), and methodological challenges (e.g., how to observe safety voice) need to be better articulated and presented.

Creating an ecological conceptualisation of safety voice

Safety voice is an ecological phenomenon. It is found to vary according to individual factors (Bienefeld & Grote, 2012), group (e.g., safety-specific transformational leadership; Conchie et al., 2012), institutions (e.g., hierarchical effects) and external environments (e.g., national culture; Malloy et al., 2009). This means that the manifestation of safety voice will vary according to the specific characteristics of a situation (individual, group, institutional, external), and corresponds to a systems approach to safety and ecological models of behaviour that specify levels of analysis (Bronfenbrenner, 1977; Erez & Gati, 2004; Leveson, 2002). It suggests that distinguishing the levels at which safety voice is analysed is important for illuminating relationships among safety voice antecedents, pragmatics and outcomes, and enabling targeted interventions (Leveson, 2002).

However, research has largely neglected conceptualising the ecological nature of safety voice, and as a consequence most scholars have not differentiated findings. Manapragada and Bruk-Lee (Manapragada & Bruk-Lee, 2016) distinguished relationship-, job-, climate- and issue-based motives for safety voice and others distinguished (work) environment antecedents (e.g., Aydon et al., 2016; Lindberg et al., 2013), but differentiation into levels of analysis is scarce. In their review of the literature, Morrow and colleagues (2016) identified eleven qualitative studies and synthesised these into four themes (i.e., hierarchies and power dynamics; perceptions of unsafe/ineffective open communication; expectations and socialisation; managerial influence), but did not offer an integrated conceptual model for relationships amongst safety voice variables, or an account of the ecological nature of safety voice. Similarly, Okuyama et al. (2014)'s review of 27 articles describing safety voice in healthcare identified the antecedents (e.g., motivation and clinical context, perceived safety of speaking-up), pragmatics (e.g., tactics involved in speaking-up), and outcomes (e.g., error correction) of safety voice. However, observations were not framed within a complete ecological conceptualisation; individual, team, and institutional factors were collapsed together, not delineated. For example, individual factors (e.g., 'roles as professionals') partly referred to social systems (i.e., organisations).

Thus, there is a need to better conceptualise the ecological nature of safety voice, and to consider research findings within this framework. Through doing this, we enable stronger interventions that can target variables across levels of analysis and reveal gaps in research. One approach would be to apply a previously established ecological model operating at varying levels of analysis to the safety voice literature, for example the hierarchical model of organisational behaviour (e.g., individual, group, organisation, external context) proposed by Erez and Gati (2004). They proposed that levels interact in top-down (e.g., organisational hierarchy enabling an individual sense of power) *and* bottom-up ways (e.g., individuals' personalities shaping teamwork), and due to this dynamic nature safety voice variables would interact and shape each other.

Optimising methodologies for investigating safety voice

Emulating the research tradition within fields such as safety culture and climate, safety voice studies have tended to utilise cross-sectional surveys (e.g., Barnett, 1992) and interviews (e.g., Aydon et al., 2016) to identify the organisational antecedents (e.g., the availability of time-outs, workload) or characteristics of those who voice safety concerns (e.g., job type, age, gender; Gauld & Horsburgh, 2014; McLaughlin, Winograd, Chung, Van de Wiele, & Martin, 2014; Nembhard, Yuan, Shabanova, & Cleary, 2015). These approaches have addressed the short-lived nature of the act of raising a safety concern (i.e., it is contained to the moment of speaking-up and difficult to encounter spontaneously) through retrospective reports by individuals who voiced or their supervisors.

Yet, the appropriateness of using surveys and interviews to investigate safety voice is uncertain. Methodologies have their strengths and weaknesses (e.g., surveys' low resource demands versus lack of depth), and the field of safety may be subject to mono-methodological bias, with surveys not yielding behavioural data, or lacking causal understanding of its drivers and outcomes. Furthermore, biases may emerge due to the nature of safety voice. For example, if safety voice involves taking a social risk to avoid physical harm, anonymised surveys and interviews may be inflated because the moral obligation to appear as a voicer may be stronger than the social risk involved in providing data.

Alternative methodologies have been used, for example interventions (e.g., Habyarimana & Jack, 2011) and observations (Bienefeld & Grote, 2012), and qualitative research has attempted to understand the pragmatics of safety voice (i.e., how one uses language to voice) and strategies to best communicate safety (Bickhoff et al., 2016; Wei Liu et al., 2016; Lyndon, 2008). Yet, the suitability of these approaches remains unclear. Thus, there is a need to assess the variety of methods to identify optimal methodologies for investigating safety voice.

The current study

To establish the relevance of the concept of safety voice it is important to investigate whether safety voice is unique in terms of its concepts, ecological nature and methodological challenges. These issues can only be addressed through systematically reviewing and synthesising the safety voice literature. Previous analyses of this literature have not addressed these concerns, and have focused only on qualitative studies (Morrow et al., 2016), or healthcare professionals (Okuyama et al., 2014). The systematic review is conceptual in nature – empirical tests of the predictive validity of safety voice in comparison to employee voice are beyond our scope – and we address three issues.

First, we address the conceptualisation of safety voice, and its overlap with employee voice. We synthesise the concepts and data from the safety voice literature into a unified framework that provides conceptual clarity. We do this in order to illuminate the conceptual uniqueness of safety voice from employee voice and reflect on its congruency with Morrison's (2011) model of employee voice, which is the only model that explicitly lists antecedents, voice and outcomes. Thus, we aim to identify the nature and usefulness of the safety voice concept, and ask: *1) what definitions, conceptualisations and theoretical backgrounds characterise safety voice?*

Second, we establish an ecological approach to conceptualising safety voice. To date, a lack of an ecological framework of safety voice has driven scholars towards limited coverage of the phenomenon, with observations being generalised or specific to narrow situations. Through reviewing the literature, the various individual, group, institutional and external variables that influence safety voice can be identified and synthesised, with an over-arching conceptual model being produced. We thus aim to evaluate the empirical evidence on safety voice to date, propose an ecological framework for future research on safety voice, and ask: *2) what is the ecological nature of safety voice in terms of inhibiting (promoting) relationships between safety voice and antecedents, pragmatics, and outcomes across levels of analysis?*

Third, we evaluate methodological approaches to investigating safety voice. Safety voice research faces unique challenges in researching a phenomenon for which harm can be immediate (i.e., imminent harm can be prevented) and proximal (i.e., it can affect victims' bodies), the social desirability of speaking-up is diffuse (i.e., it is a risk and moral obligation), and its behavioural nature may obscure post-hoc methodologies (i.e., self-reports may be inaccurate). However, the effectiveness of different methodologies has not been reviewed and challenges to researching safety voice remain unaddressed. Accordingly, we ask: *3) which* *methodologies have been used for researching safety voice, how suitable are they, and what methodological challenges remain?*

Methods

Search strategy

This review followed PRISMA guidelines for systematic reviews (Liberati et al., 2009) and applications of these (Reader et al., 2014). The search strategy for this study was pre-registered on PROSPERO (Noort et al., 2017), and is presented in Fig. 3.1. The inclusion/exclusion criteria are outlined in Table 3.1.

Identification. English peer-reviewed articles were identified through using the PubMED and EBSCOhost search engines. A preliminary search revealed that relevant articles included 'safety' and a variation of 'voice', 'speaking-up/out' or 'silence', and articles on diverse subjects (e.g., prosocial, citizenship behaviours). However, safety voice is difficult to disentangle from scales containing items on safety voice (e.g., 'I am willing to warn other coworkers about working unsafely') and other behaviours (e.g., "I am willing to pick up work-place litter that I did not cause myself"; Geller et al., 1996). Furthermore, including '*communicat*' provided an unwieldy number of hits (± 7 times more), and 'organi?ation*' specified an inappropriate a-priori contextual bound. Hence, to provide hits specifically related to safety voice (e.g., safety-related silence) the search term "(*safe* AND silenc*) OR (*safe* AND speak*) OR (*safe* AND voic*)" was used to search in titles and abstracts in the following databases: Anthropology plus, Business source complete, CINAHL Plus with full text, Communication and mass media complete, Criminal justice abstracts with full text, History of science, technology, medicine, International political science abstracts, MEDLINE, Peace research abstracts, PsycArticles, PsycINFO, SocINDEX with full text.

Table 3.1. Inclusion and exclusion criteria.

Domain	Include	Exclude
Publication	• Peer-reviewed articles	• Duplicates • Book/film/literature reviews • Periodicals • Editorials • Dissertations • On-going, unpublished trials • Errata (unless referring to extracted information) • Conference proceedings (e.g., keynotes, panel discussions)
Language	• English	• All other languages
Subject	• Raising safety concerns to another person (behaviour)	• Voice technology (e.g., VoIP) • Language skills • Physiological voice • Technology-mediated communication • Medical diseases • Primarily about ethics • Primarily about law • Intimate relationships • Drugs or therapies (not raising concerns)
Method	• Empirical, primary study • Quantitative • Qualitative	• No original data • People speak-up when prompted (e.g., focus group) • Authors advocate for a group • Calls to speak-up about an issue
Participants	• Behavioural • From/to all staff roles/hierarchies	• Non-adult participants • Mental health patients • Unions or organisations advocating an issue
Predictors	• All predictors	• No predictors discussed
Outcomes	• All outcomes	• N/A

Irrelevant articles during the preliminary search related to physiological voice (e.g., pitch), voice technology (e.g., Voice over Internet Protocol), language skills (e.g., English-speaking), and wrong participants (e.g., children, mental health patients, intimate relationships). Therefore, in the final search hits were excluded based on the following search term: “Technolog* OR VoIP OR Computer? OR PC OR Pitch OR Intonation OR Anatomy OR vocal OR Child* OR “Mental health” OR Contraception OR HIV OR molecular OR therapy OR airway OR syndrome OR “Language skill?” OR "-speaking"”. A detailed search history is published online (see supplementary material)[12].

[12] Please also see Appendix A for the review protocol (i.e., search strategy, inclusion criteria, data extraction, search development).

Screening and Eligibility. After the removal of duplicates, the title and abstracts of identified articles were screened based on the inclusion criteria (see Table 3.1). Finally, full-text articles were retrieved and checked for their eligibility. Inter-coder reliability was established through double coding 15% of the 3031 hits by a research assistant trained on applying the inclusion criteria. Gwet's AC1 was calculated as it is a robust measure for datasets with a high prevalence of one category (e.g., excluded articles; Wongpakaran et al., 2013). Very good agreement was found, indicating a reliable application of the inclusion criteria, $AC1 = .92$ (95CI: .89-.95), $p < .001$.

Data extraction and analysis[13]

Descriptive data

Data extraction. (i) author(s), (ii) year of publication, (iii) journal, (iv) country, (v) industry or context, (vi) number of studies in a publication, (vii) amount of safety voice. Quality indicators (Cochrane, open science, inter-coder reliability): (i) comparability of study groups, (ii) appropriateness of randomisation, (iii) whether randomisation was performed blindly, (iv) sample representativeness (i.e., response rate > 40%), v) appropriateness of exclusions, (vi) compliance with ethical standards, (vii) appropriateness of treatment of missing data, (viii) achievement of inter-coder reliability, and (ix) open science achievements[14]. Cochrane quality indicators are developed for randomised control trials (Higgins et al., 2011). Hence it was recorded when quality indicators were not applicable.

Data analysis. Trends on publication history and outlets, country, industry distribution, and quality indicators. Calculations for the amount of safety voice (weighted by study size, limited

[13] Supplementary data to this article can be found online at https://doi.org/10.1016/j.ssci.2019.04.039.

[14] We employed a broad definition of open science where open science was coded when follow-up research was supported (e.g., through providing materials, data or inviting opposing views). Whether publications were open access was not recorded because this is often achieved through paying a fee.

to pre-intervention amounts, aggregated across sub-groups) reported in articles. Likert scale scores were converted to percentages (e.g., 4.5 on 5-point Likert scale indicating safety voice is calculated as (4.5-1)/(5-1) = 87.5%).

Concept of safety voice

Data extraction. (i) Definition of safety voice, (ii) theoretical background. Theoretical backgrounds (i.e., theories or models) were deducted from the argument and cited literature if none was explicitly stated.

Data analysis. Trends on the theoretical background of studies, and synthesis of definitions of safety voice through coding and thematically grouping key concepts of extracted definitions (e.g., 'speaking-up about safety' was coded as: 'safety-related issues' and 'speaking-up').

The ecological nature of safety voice

Data extraction. (i) Individual-level antecedents, (ii) group-level antecedents, (iii) institutional-level antecedents, (iv) hazard-specific antecedents, (v) other antecedents, (vi) safety voice pragmatics, (vii) safety voice outcomes, (viii) direction of relationship (i.e., promoting/inhibiting), (ix) voice variable statistics. Variables were only used for the synthesis when a significant relationship (e.g., thematic, correlational, regression) was suggested in the study's results section.

Data analysis. Qualitative synthesis of the evidence on safety voice through coding and thematically grouping antecedents, pragmatics and outcomes into first and second order themes (e.g., 'fear of retaliation' and 'sense of safety' were coded as 'fear for consequences' and the higher order theme 'perceived cost of voice'), and according to level of analysis. After bottom-up synthesis, we adopted labels from Morrison's model of employee voice (Morrison, 2011) when the constructs overlapped. All grouped variables were recoded to reflect the same relationship to safety voice (i.e., promoting/inhibiting).

An ecological framework is proposed through i) providing a consistent terminology for similar but differently named antecedents and outcomes (Dixon-Woods et al., 2006; Reader et al., 2014), ii) thematically integrating related antecedents for the individual, group, institutional, external and hazard-specific levels, iii) thematically integrating related outcomes, and iv) identifying promoting or inhibiting relationships to safety voice.

Methodological data

Data extraction. (i) Operationalisations, (ii) quantitative/qualitative methods, (iii) methodology, (iv) unit of analysis, (v) independence of dataset, (vi) manipulations, (vii) interventions, (viii) high-level participant information, (ix) number of participants, (x) number of outliers deleted.

Data analysis. Trends on the use of methods and operationalisations through coding and grouping these based on similarity (e.g., structured and open-ended interviews were coded as 'interviews').

Results

Search results

A total of 48 articles met the inclusion criteria (see Table 3.1), including five articles identified through hand-search. The articles included 50 studies, with one article including three studies (Manapragada & Bruk-Lee, 2016). The extent that people raised safety concerns was provided by 24 studies ($m_{weighted}$ = 44%; SD = 4%), and an effect size could be extracted (or calculated) from 15 quantitative studies. However, 62 effect sizes were dispersed over 42 distinct variables (i.e., many variables had only up to two effect sizes). Therefore, at this stage of the literature, the planned meta-analysis would not provide additional information beyond repeating authors cited. The amount of safety voice was not associated with the number of extracted variables, or context of research (i.e., healthcare or USA versus other contexts).

Description of the safety voice literature

Interest in safety voice emerged recently (i.e., 45 of the 48 articles were published in the last decade), and publication outlets were diverse. Publications were spread over 37 journals, and the outlet containing most publications, AORN Journal, had only three publications.

Study contexts were heavily biased towards the United States (i.e., 50% of the studies). Three studies had an international sample (Anicich et al., 2015; Bienefeld & Grote, 2012; Malloy et al., 2009). Similarly, indicating the need for a context-agnostic framework, an industry-bias exists towards research in healthcare (n = 41). Further research was conducted in transport (Habyarimana & Jack, 2011; Tucker, Chmiel, Turner, Sandy Hershcovis, et al., 2008), aviation (Bienefeld & Grote, 2012), wildland firefighting (Barton & Sutcliffe, 2009), secondary education (Turner et al., 2015b), oil and gas (Conchie et al., 2012) or across industries (n = 2; Barnett, 1992; Manapragada and Bruk-lee, 2016 (Study 1)). Furthermore, despite inclusive inclusion criteria for contexts, no studies were found outside of institutions with, arguably, mountaineering groups (Anicich et al., 2015) being the least institutional in nature. These biases are important: as a proportion of extracted variables (i.e., antecedents, pragmatics, outcomes), USA-based studies have identified fewer group-based antecedents (i.e., 23% vs 31% for other countries), and healthcare-based studies identify more organisational antecedents (24% vs 8% for other industries) but less individual antecedents (32% vs 42%). This suggests context is important for researching and contextualising safety voice, and the literature might address whether unique antecedents exists across contexts (e.g., daily life).

In terms of publication quality (i.e., Cochrane, open science, inter-coder reliability), studies varied in applicable quality indicators (M = 5.86; SD = 1.41) of which they met 45% (SD = 20%). Most studies reported obtaining ethical approval or following industry-standard procedures (n = 32; e.g., obtaining informed consent). However, eighteen studies did not report sufficient information to determine ethical standards. The weakest quality of the safety voice literature is its treatment of missing data. That is, most studies did not report procedures (n = 24) or used list-wise deletion rather than data imputation (n = 4). Finally, only fifteen studies promoted open science: thirteen studies provided study materials and/or data (i.e., online, printed, or through the author), one had an invited commentary (Barzallo Salazar et al., 2014) and another a large online appendix on the study methodology (Anicich et al., 2015). Six open-science articles (38% of all articles in this period) were published before the recent growth in open science (around the year 2013 when the Centre for Open Science was founded); nine (28%) after 2013. This indicates a need to publish higher quality articles in terms of clarifying ethics, improving the treatment of missing data, and promoting open science. Descriptive information regarding the studies is presented in Table 3.2.

The concept of safety voice

Definitions

Thirty-two studies provided a definition of safety voice, and 110 concepts were identified across the definitions (M = 3.4; SD = 1.6). Some definitions were shorter (Manapragada & Bruk-Lee, 2016) than others (e.g., the definition stated above by Tucker et al., 2008). Definitions of safety voice emphasised communication (n = 30), unsafe situations (n = 36), discretion (n = 2), improvement-focus (n = 12), actors (e.g., from lower ranked staff; addressed to superiors; n = 11), and that it originates with a perception of a situation (n = 4). Five concepts were very generic (i.e., that safety voice is a motivation, willingness or ability), seven referred to variable aspects of delivering safety voice (e.g., assertiveness, persistence), and five other concepts put a stringent theoretical bound on the context of safety voice (e.g., work-related issues; Nembhard, Yuan, et al., 2015). Synthesising these concepts, safety voice may thus be defined as: explicit communication that is (1) discretionary, (2) aimed at improving a perceived unsafe situation, and (3) addressed to others of equal or senior status.

Theoretical backgrounds

Forty-eight studies referred to a total of 79 theoretical paradigms (M = 1.58; SD = .91). Most prevalent were references to safety (n = 17; e.g., patient safety, safety culture, high reliability organisations), voice (n = 16; e.g., employee voice, speaking-up decisions, patient complaints), and working in teams (n = 17; e.g., leadership, organisational culture, CRM, workplace ostracism). The remaining paradigms referred to (i) diverse theoretical or pragmatic models (n = 15; e.g., social exchange theory, social influence, social defence theory, moral courage, national cultural values, black swan theory), (ii) patients (n = 5; e.g., patient advocacy), (iii) broad areas of interest (n = 6; e.g., attitudes, communication, quality improvement), or (iv) methodology-informed paradigms (n = 2; e.g., narrative standpoint, life-world phenomenology). This highlights that the literature on safety voice is theoretically disintegrated, and that it requires (i) a unified theoretical paradigm, and (ii) a clarification of overlaps and contrasts between different theoretical accounts.

The ecological nature of safety voice

The 50 studies described 256 antecedents to safety voice. These were delineated into levels of analysis as individual (n = 94; m = 1.84, SD = 2.18), group (n = 78, m = 1.56, SD = 1.86), institution (n = 64, m = 1.28, SD = 1.59), or external context (n = 3, m = .06, SD = .31). Seventeen antecedents (m = .34, SD = .63) were hazard-specific. This highlights that an ecological safety voice framework needs to incorporate properties of hazards to account for whether people voice their safety concerns. Less research investigated the pragmatics of the safety voice act (i.e., 3 studies, n = 7, m = .14, SD = .64) or outcomes (i.e., 15 studies, n = 23, m = .45, SD = .91). To provide a consistent terminology, the 256 antecedents, 7 pragmatics and 23 outcomes were collapsed based on resemblance within their level of analysis. This left 65 first-order antecedents (e.g., impact of harm, likelihood of harm) and 31 second-order antecedents (e.g., risk). Four types of safety voice outcomes were identified (i.e., negative experiences, reduced physical harm, action-driven communication, organisational performance). Through summarising findings, this demonstrates that safety voice has been mostly conceptualised in terms of antecedents (predominantly individual), and in particular that the field has researched the phenomenon as an ecological phenomenon. All antecedents, pragmatics and outcomes of safety voice across levels of analysis, and a conceptual comparison against Morrison's (2011) model for employee voice, are presented in Table 3.3.

Table 3.2. Descriptive information of articles included in the review.

Authors	Journal	Study	Country	Setting	Methods	Out	Prag	Antecedents				Ext
								Haz	Ind	Gr	Org	
Anicich, et al. (2015)	PNAS	1	27 countries	Mountaineering	Vignettes	1	0	0	0	1	0	0
Aubin & King (2015)	Journal of Interprofessional Care	1	Canada	Healthcare (education)	Experiment	0	0	0	2	1	1	0
Aydon, et al. (2016)	Journal of Clinical Nursing	1	Australia	Healthcare (neonatal care)	Interviews	0	0	0	2	5	4	0
Barnett (1992)	Journal of Business Ethics	1	USA	Cross-industry	Survey	0	0	0	0	0	2	0
Barton & Sutcliffe (2009)	Human Relations	1	USA	Wildland Firefighting	Interviews	2	0	0	2	4	0	0
Barzallo Salazar, et al. (2014)	Journal of the American College of Surgeons	1	USA	Healthcare (education)	Experiment	1	0	0	0	1	0	0
Bickhoff, et al. (2016)	Nurse education today	1	Australia	Healthcare (education)	Interviews	0	1	0	0	0	0	0
Bienefeld & Grote (2012)	Aviation Psychology and Applied Human Factors	1	Europe	Aviation (air crew)	Observation, survey	0	0	1	4	3	2	0
Blanco et al. (2009)	AORN Journal	1	USA	Healthcare (surgery)	Survey	1	0	0	0	0	0	0
Conchie, et al. (2012)	Journal of Occupational Health Psychology	1	UK	Oil & Gas	Survey	0	0	0	3	1	0	0
Davis, et al. (2013)	Journal of Evaluation in Clinical Practice	1	UK	Healthcare (surgery)	Focus groups	0	0	0	0	0	1	0
Delisle, et al. (2016)	Journal of Interprofessional Care	1	Canada	Healthcare	Survey	0	0	0	0	0	0	0
Frosch, et al. (2012)	Health Affairs	1	USA	Healthcare (primary care)	Focus groups	0	0	0	1	3	3	0

Gauld & Horsburgh (2014)	Australian Health Review	1	New Zealand	Healthcare	Survey	0	0	0	2	0	1	0
Gkorezis, et al. (2016)	Journal of advanced nursing	1	Cyprus	Healthcare (nursing)	Survey	0	0	0	4	1	0	0
Gurung, et al. (2017)	BMC health services research	1	Nepal	Healthcare	Interviews	0	0	1	2	2	4	0
Habyarimana & Jack (2011)	Journal of Public Economics	1	Kenia	Transport (long-distance road transportation)	Intervention	2	0	0	0	1	0	0
Hanson (2017)	Journal of infusion nursing	1	USA	Healthcare (intensive care)	Survey, quasi-experiment	0	0	0	0	0	0	0
Hemingway, et al (2015)	AORN Journal	1	USA	Healthcare	Survey	0	0	0	0	0	0	2
Henkin, et al. (2016)	Journal of multidisciplinary healthcare	1	USA	Healthcare (general medicine)	Intervention, survey	0	0	0	1	0	0	0
Hovey, et al. (2011)	Qualitative health research	1	USA	Healthcare	Interviews	1	0	0	1	2	1	0
Howard, et al. (2013)	Journal of patient safety	1	Australia	Healthcare (acute care)	Interviews	0	0	0	2	0	0	0
Hu, et al. (2015)	Journal of the American College of Surgeons	1	USA	Healthcare (surgery, anaesthesiology, nursing)	Observation	0	0	0	0	1	0	0
Jackson, et al. (2010)	Journal of Advanced Nursing	1	Australia	Healthcare (nursing)	Interviews	5	0	0	3	0	2	0
Jacobsson, et al. (2012)	Scandinavian journal of trauma, resuscitation and emergency medicine	1	Sweden	Healthcare (trauma)	Observations	0	0	0	0	2	0	0
H. L. Johnson & Kimsey (2012)	AORN Journal	1	USA	Healthcare (anaesthesia, surgery, obstetrics, gynaecology,	Survey	1	0	0	0	0	0	0

				perioperative services)								
K. R. Johnson, et al. (2006)	Archives of dermatology	1	USA	Healthcare (cross-sectional)	Survey	0	0	0	2	0	0	0
A. Jones, et al. (2016)	International Journal of Nursing Studies	1	England, Wales (UK)	Healthcare (executive boards)	Interviews	0	0	0	0	6	1	0
Kulig & Blanchard (2016)	Journal of graduate medical education	1	USA	Healthcare (anaesthesia)	Intervention	1	0	0	9	1	0	0
Lindberg, et al. (2013)	International journal of qualitative studies on health and well-being	1	Sweden	Healthcare (elderly care nurses)	Interviews	0	0	1	2	8	5	0
Liu, et al. (2016)	Journal of clinical nursing	1	Australia	Healthcare (acute care)	Critical ethnography	0	4	0	0	3	1	0
Lyndon (2008)	Journal of Obstetric, Gynecologic, and Neonatal Nursing	1	USA	Healthcare (perinatal care)	Interviews, observation	2	2	1	4	3	4	0
Malloy et al. (2009)	Nursing Ethics	1	Canada, Ireland, Australia, Korea	Healthcare (nursing)	Focus groups	0	0	0	1	0	1	1
Malvey, et al. (2013)	Journal of Healthcare Management	1	USA	Healthcare	Survey	0	0	2	2	1	2	0
Manapragada, et al. (Manapragada & Bruk-Lee, 2016)	Accident Analysis and Prevention	1	USA	Safety-critical organisations	Focus groups, interviews	0	0	1	1	2	2	0
		2	USA	Healthcare (Nursing)	Survey	0	0	1	0	1	2	0
		3	USA	Healthcare (Nursing)	Survey	1	0	1	3	1	3	0
Manias (2015)	Health Expectations	1	Australia	Healthcare (surgery, medical teaching)	Interviews	1	0	1	5	3	0	0
Maxfield, et al. (2013)	American Journal of Obstetrics & Gynaecology	1	USA	Healthcare (labour-and-delivery)	Survey	0	0	0	2	2	0	0

McLaughlin, et al. (2014)	World neurosurgery	1	USA	Healthcare (surgery)	Survey	0	0	0	0	0	2	0
Nembhard et al. (2015)	Health Care Management Review	1	USA	Healthcare (primary care)	Survey	2	0	0	0	0	1	0
Nembhard, et al. (2015)	Health Care Management Review	1	USA	Healthcare	Interviews	0	0	0	8	0	7	0
Noland & Carmack (2015)	Qualitative health research	1	USA	Healthcare (nursing education)	Interviews	0	0	2	3	1	1	0
Phelps & Reed (2016)	Canadian Journal of Infection Control	1	USA	Healthcare	Focus groups	0	0	0	6	4	1	0
Schwappach & Gehring (2014c)	PLOS ONE	1	Switzerland	Healthcare (oncology)	Vignette study, survey	0	0	2	6	5	4	0
Schwappach & Gehring (2014d)	BMC Health Services Research	1	Switzerland	Healthcare (Oncology)	Interviews	0	0	2	2	0	3	0
Schwappach & Gehring (2014b)	BMJ Open	1	Switzerland	Healthcare (oncology)	Interviews	0	0	1	0	4	1	0
Seiden, et al. (2006)	Quality & Safety in Health Care	1	USA	Healthcare (education)	Case study	1	0	0	1	4	1	0
Tucker et al. (2008)	Journal of Occupational Health Psychology	1	UK	Transport (bus drivers)	Survey	0	0	0	5	1	1	0
Turner, et al. (2015b)	Journal of Safety Research	1	Canada	Education (high schools)	Survey	1	0	0	1	0	0	0

Abbreviations: Out.: Outcome; Prag.: Pragmatics; Haz.: Hazard; Ind.: Individual; Gr.: Group; Org.: Organisational: Ext.: External context

Methodological challenges

In terms of datasets, 39 studies had an independent dataset. Nine studies were part of a larger project, and two included reinvestigated data (Schwappach & Gehring, 2014b, 2014c). A total of at least 52,948 participants took part across the studies (M = 1177; SD = 3490; median = 135). The exact number of participants could not be determined for four studies (Hanson, 2017; Henkin et al., 2016; Phelps & Reed, 2016; Seiden et al., 2006), and twelve studies removed data (e.g., attrition, partial data).

Studies were qualitative (n = 23), quantitative (n = 22), or mixed-method (n = 5). Methodologies included surveys (n = 17), interviews (n = 12), focus groups (n = 4), field or lab experiments (n = 2), vignette studies (n = 2), video observations (n = 2), case studies (n = 1), and combinations of these (n = 8). Quasi-experimental methods for studying safety voice were limited (n = 1; Delisle et al., 2016), and few studies manipulated safety voice (n = 5, i.e., 2 through leader behaviours, 2 vignettes, 1 intervention timing) or tested safety voice interventions (n = 7, i.e., 4 training programs, 3 changes to the environment).

Safety voice variables (i.e., antecedents, pragmatics, outcomes) across levels of analysis were identified through quantitative (n = 67) and qualitative methods (n = 183), but qualitative studies identified relatively more group-based antecedents (32% vs 22% of all identified variables) and fewer outcomes (7% vs 15%).

To elicit data on safety voice, 32 studies used a single operationalisation of safety voice and eighteen studies used two (e.g., latent themes and scales). Most studies relied on safety voice as identified through latent themes (n = 28) or scales (i.e., targeted at individuals, n = 16, groups, n = 3, or institutions, n = 2).

Situational realism was approached through episode-recall (n = 11), vignettes (n = 2), or through in-situ observations (n = 5; e.g., observation or text analysis of transcribed conversations). Finally, only seventeen studies explicitly operationalised the victim in safety voice (whether self or other). Post-hoc reports were a dominant method to elicit data from voicing individuals (n = 40) or their seniors (n = 4). The use of external sources (e.g., databases, behavioural observation, insurance claims) was rare (n = 6).

Together, this indicates that 38 safety voice studies (76%) base findings on associations (i.e., correlations or themes). Research using broad-ranging methodologies has identified numerous variables and it seems appropriate to engage in a new phase of experimental research that can support the development and assessment of interventions in-situ. Furthermore, a need exists for improving safety voice operationalisations through (i) relying less on post-hoc reports that might bias results, (ii) disentangling victimhood from voice, and (iii) increasing the realism of safety voice scenarios.

Table 3.3. The Safety Voice framework.

Hazard		+	-	n	**Individual**		+	-	n	**Group**		+	-	n	**Institution**		+	-	n	**External context**		+	-	n
1. Antecedents																								
Risk	High impact of harm	7		7	Demographics*	Older age	5	1	6	Norms	Unfavourable norms		5	5	Structures*	Larger size	1		1	National culture	Confucian		1	1
	High likelihood of harm	2		2		Female gender	4		4	Access	Legitimated in concerns	4		4		Segregated organisation		2	2	Outside interest	Media coverage	1		1
Speed	High speed		2	2		Higher SES	1		1		Ostracism	1	2	3		Group membership	n/a		7	Regula-tion	National guidelines	1		1
Awareness	Hazard noticed	2		2		Longer tenure and experience	3		3		Territorial colleagues		1	1		Hierarchy		8	8					
Information	Unambiguous situation	1		1	Perceived cost of voice*	Fear for consequences	2	11	13	Situational presence	Bystanders		2	2		Enabling structures/ resources	6		6					
	Involves personal information		1	1		Need to protect reputation		7	7		Self	1		1		Lack of open communication channels		3	3					
Controllability	Within control		1	1		Trusting others	4		4	(Leadership) Openness*	Receptive others	12		12	Culture*	Unfavourable culture		9	9					
Trade-offs	No conflicting motives		1	1		Favourable previous experiences	5		5		Invited or encouraged	9		9		Support	5		5					
					Personality*	Openness to experience	1		1	Relationship concerns	Good relationship with receiver*	7		7	Information	Legal advice against voice		1	1					
						Conscien-tiousness	1		1		Fragile relationships*		8	8		Availability of information	3		3					
						Extraversion	1		1		Safety voice affects others		2	2	Work configuration	High workload	1	14	15					
						Agreeablenes	1		1	Hierarchy and leaders' actions*	Leadership style: consistent and transformational	3		3		Classroom-reality mismatch		1	1					
					Motivation*	To protect others	3		3							Busy work environment	1	2	3					
						Willingness to voice	3		3		Group hierarchy		7	7										
						Felt Responsibil-ity	7		7		Difficulties due to authority		3	3										
					Position and status (power)*	Seniority	6		6	Shared safety knowledge	Other's are responsible/ know		3	3										
					Safety knowledge	Subject expertise	9	1	10		Others are fallible	4		4										

					Intuition to speak-up	1		1		Shared understanding*	2		2				
				Perceived efficacy*	Ability to speak-up	6		6	Opportunities	Learning opportunities	1		1				
					Self-confidence	5		5	Receiver type*	Community members	1		1				
				Perceived opportunity*	Sense of futility		5	5									
				Identification	With the organisation	1		1									
2. Safety voice pragmatics																	
				Language use*	Redefine issue			1									
					Work the hierarchy			1									
					Politeness			1									
					Increment urgency			1									
					Frame as support			1									
					Mirror receiver's language		2										
3. Outcomes																	
Physical harm*	12		12	Negative experiences*	Negative responses*	3		3	Action-driven communi-cation*		3		3	Performance*	3	1	4
					Sense of failure*		1	1									

+ Promoting/promoted by Safety Voice; - Inhibiting/inhibited by Safety Voice; * Overlaps with Morrison's (2011) model for Employee Voice.
Note: where appropriate, extracted antecedents were recoded to provide a uniform polarity (e.g., all age antecedents were recoded towards indicating high age). A smaller font is used to enable all columns to be fitted onto one page.

Discussion

This study systematically reviewed the safety voice literature, which consisted of 48 articles, most of which have been undertaken in healthcare and the US. In comparison to the safety voice model of Okuyama and colleagues (2014) our framework provides a more comprehensive overview (i.e., 65 versus 19 antecedents; 6 versus 3 pragmatics; 4 versus 2 outcomes) across clearly delineated levels of analysis. Three questions were addressed, and we revealed the state of the safety voice concept (i.e., grounded in diverse theories, definitions and conceptualisations), its ecological nature (i.e., spread across hazard, individual, group, institutional and external levels of analysis) and methodological challenges (i.e., high rate of research using report-based methods and contexts of the USA, healthcare). We revealed for the first-time the safety voice outcomes, pragmatics and antecedents that inhibit (or promote) safety voice at Erez and Gati's (2004) levels of analysis for the individual, group, institutional and external context.

These findings have challenges and implications for safety voice, and we expanded upon these in the following sections.

The conceptual uniqueness of safety voice

Our systematic review revealed that safety voice had important conceptual overlaps with employee voice, yet also unique aspects that only warrant partial conceptual integration.

Behaviourally, safety voice appeared similar to employee voice. That is, safety voice concepts and definitions, like employee voice, described a verbal behaviour in which people communicate a concern to others (e.g., colleagues) to change a perceived situation, with a similar propensity (i.e., discrete, constructive, proactive), and grounded in communication and teamwork concepts (Morrison, 2011). Furthermore, whilst future work may uncover this, we did not identify studies describing unique variables for social risk, moral obligation or proximity of outcomes. This suggests that safety voice and employee voice may be difficult to distinguish in practice. For example, within safety-critical environments, concerns regarding the lack of protective equipment during a procedure can involve a safety issue (e.g., the new procedure can cause harm), a work-related issue (e.g., leadership does not comply with its responsibility to provide protective equipment), or bullying (e.g., being forced to work under lower standards than colleagues). This overlap highlights the potential for integrating behavioural concepts on speaking-up (Wilkinson et al., 2019), and a need for research to investigate the extent of empirical overlap in practice. We did not present evidence for this, and we anticipate future empirical investigations will prove fruitful.

However, safety voice appeared unique in terms of the content of the raised message, the context and person speaking-up, and identified antecedents. That is, because safety voice involves raising a safety concern in response to a perceived hazard it appears closer to prohibitive (i.e., concerns about practices that may harm organisations) than promotive employee voice (Liang et al., 2012). This is important because it suggests safety voice's scope may be limited to prohibitive messages (i.e., preventing harmful outcomes), and the safety content may provide a unique type of message and voice behaviour (Morrison, 2011). The issue of risk perception appears important for distinguishing safety voice: the need for safety voice hinges on the perception of a safety problem, and the recognition that it requires addressing. We agree with Morrison and colleagues (Morrison, 2011; Wilkinson et al., 2019) that voice research (in the broadest sense) should emphasise the content of the communicated messages, whilst continually evaluating potential for integration across voice types.

Furthermore, safety voice extends beyond organisational environments. Safety concerns are raised by other persons than employees (e.g., patients, family members, friends, bystanders; Hu et al., 2015), and beyond organisational contexts (e.g., in public, during sports activities; Anicich et al., 2015). Part of the unique value of safety voice thus resides in the broader context of hazardous situations and future research should explore hazardous scenarios outside of organisational environments.

Finally, in comparison to Morrison's model for employee voice (Morrison, 2011), unique antecedents exist for employee voice (e.g., job attitudes, learning, decision-making, group harmony) and safety voice (i.e., safety knowledge, norms, shared safety knowledge, information, work configuration, national culture, outside interest, regulation), with hazard-related antecedents revealing the unique scope of safety voice (see Table 3.3). Arguably, this reveals a difference in the fields' aims and scopes, and future research may identify further overlaps in antecedents.

Assumptions in researching safety voice

At least three distinctive assumptions appear to shape the safety voice literature in terms of theory and methods. These relate to the social nature of processes involved in safety voice, the cost and benefits of safety voice outcomes and the research context.

First, meaning is attributed to absences, and this implies that safety voice is a process of social construction. That is, silence (i.e., the absence of voice), and in particular safety (i.e., the absence of physical harm), are considered to constitute relevant concepts with real implications. This is important because things that have *not* occurred are difficult to assess, and manage, without invoking factors that precede, replace or follow from the absence, and it is difficult to understand safety voice without taking into account how process of social construction can create opposing views. For example, a lack of safety voice may only be meaningful because beliefs on the nature of physical harm, the desirability of outcomes, and absent behaviours are socially constructed (Lupton, 1999; Turner & Gray, 2009). Yet, what is considered to be safe or desirable may be ambiguous, contested, and altered over time through sense-making processes, and this suggests that future research needs to address the act of raising safety concerns as inherently meaningful, social and embedded in a sociocultural context (Weick, 1995, 2010).

Second, like employee voice (Bashshur & Oc, 2014), safety voice is treated as producing mainly favourable outcomes. However, physical harm may carry a larger cost than unresolved work issues, and this may imply that people would rationally consider it optimal to raise more safety than work issues. This review highlighted that research investigating the complete safety voice process (antecedents, voice pragmatics, output) are scarce, and when outcomes were included these focussed predominantly on prevented physical harm (e.g., preventing wrong site surgery; Blanco et al., 2009). Yet, safety voice outcomes may also be unfavourable (e.g., negative experiences), mixed or ambiguous and the 'expected utility calculus' (Milliken et al., 2003) predicts this would reduce the likelihood of people speaking-up. This distinguishes safety voice from employee voice, because this would imply that if each safety voice instance has a cost (e.g., negative responses from others) and each prevented safety incident a significantly larger benefit (e.g., physical harm), then, dependent on the ambiguity of the expected utility of the outcomes, individuals would rationally produce a large number of false alerts for each correct alert in the case of safety voice. Approaches from signal detection theory (Nesse, 2005) and game theory (e.g., J. S. Brown et al., 1999) have been applied to defensive responses such as fear, and utilising these approaches might prove useful to uncover costs and benefits of safety voice outcomes and potential optimum levels of safety voice.

Finally, based on the lack of research outside of organisational and clinical contexts, the literature seems to have assumed that safety voice is exclusively an organisational phenomenon. This finding highlights and important gap in safety voice research to date: safety risks are not confined to work contexts (Wilson, 1979) and people are concerned about risks that extend well beyond them (e.g., natural hazards, fires; G. W. Fischer et al., 1991). Research using experience sampling methods has highlighted that only 29% of people's concerns about physical risks are related to work activities (Hogarth et al., 2007), and concerns relate more often to personal transportation and food safety. Underscoring this, between 1979 and 2014 a total of 288,211 deaths were attributed to non-work related accidents (e.g., motor accidents, drowning, falls, poisoning) in the UK (ONS, 2016). The safety voice literature's emphasis on institutions may have emerged from (i) organisations' desire to manage and control safety and (ii) researchers' desire to use a combination of relevant and accessible data. However, to fully understand the processes involved in voicing safety concerns, the literature needs to expand beyond organisational contexts.

An ecological framework for safety voice

We set out to uncover the inhibiting and promoting relationships between antecedents, safety voice, and its outcomes across levels of analysis (Bronfenbrenner, 1977; Erez & Gati, 2004; Leveson, 2002). To this end, we proposed an ecological framework for safety voice antecedent and outcomes (see Table 3.3) that makes three contributions.

First, the framework highlights the ecological nature of safety voice for antecedents, pragmatics and outcomes. Safety voice can be delineated at Erez and Gati's (2004) individual, group, institution and external context levels of analysis, and some variables manifest differently across these levels. For example, power differentials are manifested as individual 'power', group 'hierarchy and leaders' actions', and institutional 'hierarchical structures'. Furthermore, we found that several antecedents related specifically to characteristics of hazards. This is important for safety voice as an ecological phenomenon, because it implies that in addition to outcomes, the unsafe event dynamically shapes and is shaped (i.e., a feedback loop) by the social context of individuals, groups, institutions and external environment. The framework therefore enables a systems approach to safety voice (Leveson, 2002), that is not prescriptive but describes the nature of relationships amongst antecedents, voice pragmatics and outcomes across levels of analysis as dynamic and emergent (White, 1995).

Second, the framework suggests novel directions for research. The framework simultaneously reveals gaps and abundances in knowledge. Arguably, certain domains would not require additional evidence (e.g., workload, receptiveness of others, physical harm, fear for consequences), whereas others would (e.g., unfavourable outcomes, the external context, unconscious processes, characteristics of the hazard). In particular, the literature has put a significant emphasis on antecedents, but a gap remains in terms of a clearer operationalisation of victimhood (i.e., who suffers the physical harm), safety responsibilities (i.e., people who are tasked with managing the safety issue may not speak-up), beliefs on hazard characteristics, safety voice pragmatics, and unfavourable outcomes.

Finally, the framework enables improved safety management. It facilitates causal factor analysis for accident investigation (through enabling the identification of causes for silence), and, in particular, the application of evidence-based interventions by ensuring interventions cover the available empirical evidence and identifying suitable loci for novel interventions. For example, based on evidence across levels of analysis, interventions may target the removal of trade-offs, creating (shared) safety knowledge, or providing favourable work configurations and regulation.

Methodological issues in researching safety voice

Safety voice research is characterised by broad-ranging methods (i.e., quantitative and qualitative), a reliance on post-hoc reports and a need for causal conclusions. These challenges are similar to those outlined by Morrison (2011), yet four methodological challenges require special attention to further the safety voice literature.

First, the field tends to assume that people can report on the antecedents and outcomes of safety voice, and has treated post-hoc reports as reliable data. Post-hoc reports are valuable as they are relatively resource efficient and can enable large sample sizes (e.g., through surveys). However, Podsakoff and Organ (1986) pointed out common method variance, motives for consistency and social desirability, and a range of perceptual biases (e.g., confirmation bias, illusory correlations) can bias post-hoc reports. Furthermore, self-reports may be biased through motivations to mitigate social risks (e.g., negative consequences), and supervisor-reports may be inflated through a desire to appear as an effective leader. Also, such findings are often correlational, and the causality of relationships between safety voice and its antecedents and outcomes requires explicit testing.

Second, related to this, data on safety voice occurring in-situ is scarce, and few studies have observed the complete safety voice phenomenon (i.e., antecedents, pragmatics, outcomes). Collecting data in dangerous scenarios (e.g., through exposing or prolonging participants to danger) or operationalising victimhood is ethically challenging (Lee-Treweek & Linkogle, 2000), and the short-lived nature of safety voice means it is difficult to encounter spontaneously. Report-based methodologies (e.g., statements provided in surveys, interviews, etc.) enable the circumvention of this, yet are limited as they do not observe safety voice behaviours (e.g., reports of behaviour are not the behaviour), rely on memory and imagination (e.g., poor memory and social desirability can introduce bias), and cannot establish causal relationships with other antecedent and outcomes (e.g., due to common method bias). Our review revealed limited methodological breadth, and the limitations of report-based methodologies indicates new methodologies are needed to investigate safety voice, for example experience sampling methodologies, text analysis of transcribed conversations on safety, or designing safety voice scenarios for simulations or laboratory settings. This may also address (i) the empirical relationship between safety voice and silence, which Morrison (2011) suggested is outstanding for employee voice, and (ii) the challenge of manipulating risk perceptions without crossing ethical boundaries. Thus, to address this, we investigated the challenges for investigating safety voice, and presented a novel safety voice experiment for use in laboratory settings (Noort et al., 2019a).

Third, it remains unclear whether safety voice processes are particular to their industrial or national contexts, and no data on measurement equivalence exists for safety voice in terms of its relevance, construct, and relationship with antecedents and outcomes (Reader et al., 2015). Thus, there is a need to compare results across contexts.

Finally, a need exists to clarify ethical procedures, treatment of missing data, and promote open science. This may be achieved through providing more information during the publication process, and authors need to address whether data can be shared on open platforms.

Limitations

Several limitations of this systematic review need to be stated. First, additional articles may have been published during the process of writing this article (March 2017 onwards) and the date of publication. A mere product of the academic publication cycle, authors of future safety voice reviews should take this into account. Second, systematic reviews are limited to their search strategy. For this article, the search strategy was limited to include articles when '*safe*' and not 'technolog* were mentioned in the title or abstract. This means that the search strategy may have missed employee voice studies that covered safety but did not mention this in the title or abstract, or voiced concerns through technological means. The initial searches did not reveal any articles in this regard, and we aimed to mitigate this through a hand search. Third, the strength of the safety voice framework is only as strong as the quality of the evidence that underpins it. The quality of articles and associated risk of bias has been addressed in this review, yet the framework may be biased to the extent that the field's assumptions and methods are systematically biased. We addressed above how the field can address this with future research. Finally, this systematic review had a conceptual scope and a need remains for direct empirical comparisons of safety and employee voice to establish their relative predictive validity. This is especially relevant for organisational environments where reducing harm is the main organisational goal (e.g., for air traffic control, healthcare): studies on employee and safety voice may render identical (or highly correlated) findings where the concepts' scopes overlap (i.e., reducing harm through speaking-up in safety-critical organisations).

Conclusions

Despite some conceptual and methodological overlap between safety voice and employee voice, we conclude that it is a related but distinct phenomenon due to its unique focus on safety and risk perception, the wider range of stakeholders involved, and the distinct methodologies, antecedents, and outcomes of safety voice. Understanding safety voice as an ecological phenomenon, whereby it can be influenced by hazard, individual, group, organisational, and contextual factors, can help researchers to improve understanding on why people do or do not engage in safety voice, and can support practitioners in developing interventions to enable speaking-up within specific safety-critical situations.

CHAPTER 4:
METHODOLOGICAL CHALLENGES FOR INVESTIGATING SAFETY VOICE

Preface

In the previous chapter I revealed that the safety voice literature relies on evidence from post-hoc and report-based methodologies (e.g., interviews, surveys, focus-groups) that are assumed to reliably reflect the acts of raising and withholding safety concerns. Scholars have acknowledged the implications of this assumption for the interpretation of findings (e.g., Schwappach & Gehring, 2014c). However, this means that we have not directly studied safety voice behaviour but people's perspectives on safety voice behaviour. Because of this empirical gap, it remains unclear how people actually engage in safety voice and safety silence, or how interventions alter the behaviour. Yet, whilst the broader methodological literature has identified applicable methodological and ethical challenges (e.g., Gleibs, 2014; Mitchell, 2012; Nestel et al., 2017; also see Chapter 2), it remains unclear how these apply to assessing safety voice behaviour (and safety-related behaviours more generally), and in order to better understand how safety silence contributes to accidents, a need exists for methods that can directly access safety voice alongside the extent to which people are concerned.

In Article 2, I present three studies to develop, evaluate and make accessible an experimental paradigm for investigating safety voice behaviour (termed 'Walking the Plank'). The article aimed to build an experimental paradigm for investigating safety voice, but it also creates a template for the design of new experimental paradigms in the safety management literature (e.g., for targeting other safety-related behaviours in healthcare or flight simulations, also see Chapter 7). This is important because, whilst simulation studies have been published in medical journals (Barzallo Salazar et al., 2014; Kolbe, Weiss, et al., 2013) and journals interested in organisational behaviour (Weiss et al., 2017, 2018), the top safety management journals with a generic contextual scope (e.g., Safety Science, Accident Analysis and Prevention, Risk Analysis, Human Factors) tend to publish relatively few laboratory studies[15]. Article 2 therefore aims to highlight how experiments may establish insights into safety management by addressing key methodological and ethical issues for observing safety-related behaviours (e.g., assessing risk perception, providing ethical protocols, addressing the association between reported and observed behaviour).

In terms of the mapping to the thesis research questions, these studies were originally intended to investigate how interventions are optimally tailored to the nature of safety voice (research question 3). However, because the paradigm was not yet stable it functioned more optimally to evaluate challenges to investigating safety voice behaviour (therefore contributing to research question 2: *What is the optimal way to investigate safety voice behaviour?*). In addition, I evaluate which aspects of safety voice need to be established by experiments (therefore contributing to research question 1: *What is the behavioural nature of safety voice?*) and how studies can establish causal conclusions on safety voice (therefore contributing to research question 3: *To what extent do interventions for promoting safety voice and reducing safety silence need to be tailored to the behavioural nature of safety voice?*).

To avoid being misread, I should explicitly note that I do not debate that post-hoc or report-based methods have distinct value. Studies using these types of methods have contributed important insights into the factors that shape safety voice (Morrow et al., 2016), and I have personally used report-based methods to great satisfaction in safety management studies, and beyond, because they enable in-depth insights into people's worldviews (e.g., interviews;

---

[15] A crude analysis indicates that the term 'laboratory experiment' only provides thirty hits on Google Scholar across publications in these four journals in 2019, of which ten only cited laboratory experiments. Most laboratory studies were published in Human Factors (n = 16).

Franks et al., 2017) and cost-effective data collection from large samples (e.g., surveys; Noort et al., 2016; Reader et al., 2015). Yet, methods have inherent limitations (e.g., archival studies may capture trends that have since changed, Gergen, 1973; experiments have debated external validity; Gigerenzer, 1984) and it remains unclear how these apply to observing safety voice behaviour. In addition to experiments, other methods can enable direct access to safety voice behaviour and the conclusions from this chapter (e.g., the need to establish safety concerns to establish safety silence), arguably, apply to the investigation of safety voice behaviour through simulations (e.g., Raemer et al., 2016) and archival studies that capture field-based data (see Chapter 6).

Study description

Across three studies, I evaluate how the proposed 'Walking the Plank' paradigm can address challenges to investigating safety voice behaviour. I review safety voice methods and identify nine methodological shortfalls in the literature that limit behavioural conclusions on safety voice (see Table 4.1). These relate to the literature establishing scant data on the behavioural nature of safety voice (e.g., because reports of safety voice are not safety voice behaviour), relying on recalled and hypothesised behaviour (e.g., providing potentially biased data) and providing limited insights into the relationships of safety voice with other variables (e.g., in terms of causality).

The three studies evaluate how these challenges can be addressed by laboratory experiments. Lessons were drawn on the concept (i.e., the behaviour is contingent upon risk perception, it manifests as a varied and continuous variable) and operationalisation of safety voice (i.e., the behaviour is observable, but operationalising safety silence is challenging). This indicated that in order to establish safety silence, behavioural investigations need to assess the extent to which individuals are concerned upon encountering threats, and that safety voice can

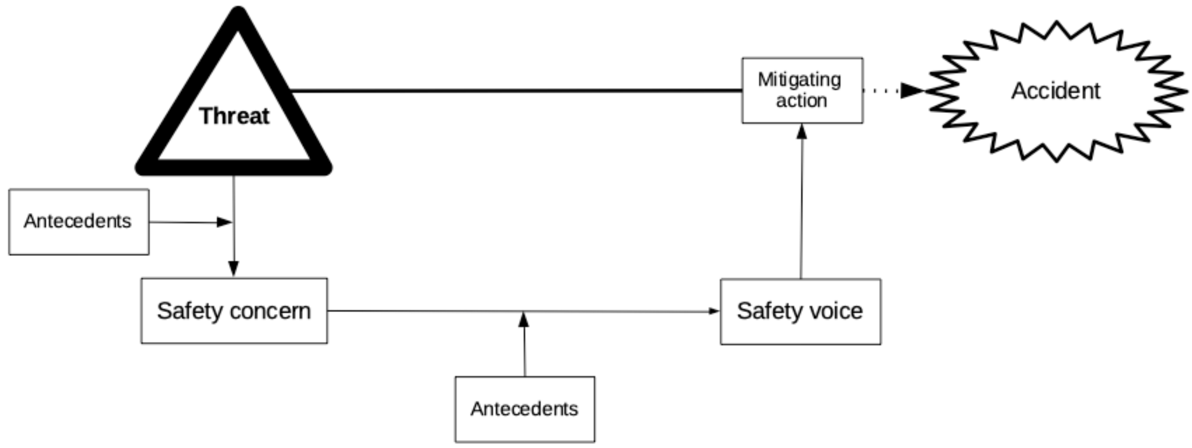


Figure 4.1. Threat Mitigation Model of safety voice and safety concerns.
Note: this model highlights that the dysfunctional momentum of threats towards accidents (Barton & Sutcliff, 2009) can be mitigated through safety voice when individuals are concerned. This emphasises that safety voice can be understood as (un)concerned voice and silence (also see Table 4.7), with antecedents reducing the extent that threats elicit concerns and voice.

be conceptualised as a two-dimensional behaviour with the extent to which hazards elicit concerns providing an important step in the prevention of harm (see Figure 4.1). Most importantly, by establishing and evaluating the Walking the Plank paradigm, I contribute a new way to investigate safety voice. This is evidenced by the successful adoption of the experimental paradigm by five MSc students for their thesis research, to date[16].

Ethics approval for the studies and supporting pilot studies was obtained from the Research Ethics Committee at the London School of Economics (see appendix B). To support future research in applying and tailoring the paradigm, I provided a protocol manual and illustrative video as online supplementary materials to the publication (Noort et al., 2019a)[17]. Copies of the protocol manual and paper-based study materials are provided in Appendix C and

[16] This involved: Eleanor McSweeney, Grace Rahal, Kate Sitniewski and Lucy Zhong (on disinhibition, gender, closed leadership), and Alyssa Pandolfo (on social norms).
[17] The online supplementary materials can be found at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00668/full#supplementary-material.

Appendix D, respectively. Readers are strongly encouraged to familiarise themselves with the illustrative video.

Authorship

"[I] was responsible for designing and conceptualising the study, data collection, formal data analysis and interpretation, and manuscript preparation (80%). [Dr Reader] and [Dr Gillespie] contributed to the conceptualization of the study, refining the study design, interpreting results, and reviewing and editing the manuscript (20%)" (Noort et al., 2019a; p.16). Research assistants aided with the execution of the protocol: Lindsie Arthur-Hulme, Arben Islami, Sai Kalvapalle, Sidsel Larsen, EleanorMcSweeney, Grace Rahal, Leia Rusch, Alyssa Pandolfo, Kate Sitniewski, Rachel Taylor, Ophelia Wong, and Lucy Zhong.

Article 2:
Walking the plank – An experimental paradigm to investigate safety voice

Published under a Creative Commons License (CC-BY 4.0) in Frontiers in Psychology – Organizational Psychology section as:

Noort, M. C., Reader, T. W., & Gillespie, A. (2019). Walking the plank: an experimental paradigm to investigate safety voice. *Frontiers in psychology, 10*, 668. https://doi.org/10.3389/fpsyg.2019.00668.

Abstract

The investigation of people raising or withholding safety concerns, termed safety voice, has relied on report-based methodologies, with few experiments. Generalisable findings have been limited because: the behavioural nature of safety voice is rarely operationalised; the reliance on memory and imagination has well-established biases; and determining causality requires experimentation. Across three studies, we introduce, evaluate and make available the first experimental paradigm for studying safety voice: the 'Walking the plank' paradigm. This paradigm presents participants with an apparent hazard (walking across a weak wooden plank) to elicit safety voice behaviours, and it addresses the methodological shortfalls of report-based methodologies. Study 1 (n = 129) demonstrated that the paradigm can elicit observable safety voice behaviours in a safe, controlled and randomised laboratory environment. Study 2 (n = 69) indicated it is possible to elicit safety silence for a single hazard when safety concerns are assessed and alternative ways to address the hazard are absent. Study 3 (n = 75) revealed that manipulating risk perceptions results in changes to safety voice behaviours. We propose a distinction between two independent dimensions (concerned-unconcerned and voice-silence) which yields a 2x2 safety voice typology. Demonstrating the need for experimental investigations of safety voice, the results found a consistent mismatch between self-reported and observed safety voice. The discussion examines insights on conceptualising and operationalising safety voice behaviours in relationship to safety concerns, and suggests new areas for research: replicating empirical studies, understanding the behavioural nature of safety voice, clarifying the personal relevance of physical harm, and integrating safety voice with other harm-prevention behaviours. Our article adds to the conceptual strength of the safety voice literature and provides a methodology and typology for experimentally examining people raising safety concerns.

Keywords: Safety Voice; Safety Silence; Safety Concerns; Experimental methodologies; Behavioural Observations

Introduction

The term safety voice describes the behaviour of raising, or withholding, safety concerns to prevent physical harm from hazardous situations (e.g., Tucker, Chmiel, Turner, Sandy Hershcovis, et al., 2008). Across organisational (e.g., healthcare, energy), family (e.g., transport, DIY), and leisure contexts (e.g., high risk sports), promoting the act of raising of safety concerns can reduce people's exposure to hazards (e.g., medicine dispensation, dangerous driving, high-altitude climbing without proper gear) and prevent physical harm (Anicich et al., 2015; Manias, 2015). The absence of speaking-up, also termed safety silence, (Barzallo Salazar et al., 2014), has been implicated in catastrophes such as the 1983 Challenger disaster (Moorhead et al., 1991) and 2010 Deepwater horizon oil spill (Reader & O'Connor, 2014), and is estimated to be involved in 25% of aviation accidents (Bienefeld & Grote, 2012; Tarnow, 1999).

Due to the difficulty of observing safety voice in safety-critical situations, academic safety voice publications tend to present data obtained through report-based data (e.g., surveys, focus groups, interviews, vignettes; Noort et al., 2019b) in which individuals or their seniors report on behavioural responses to previously held or imagined safety concerns. Yet, it remains unclear whether data from reports is reflective, explanatory, and predictive of safety voice behaviours. Alternative approaches are required to study the conditions and ways through which people raise or withhold safety concerns, and to address this, we propose and test the first experimental paradigm for investigating safety voice. Through investigating the occurrence of safety voice behaviours in a laboratory setting, and the challenges in assessing these, we aim to establish a methodology for i) observing the behavioural nature of safety voice; ii) reducing the methodological reliance on memory and imagination; and iii) advancing knowledge on the factors that predict safety voice.

Safety voice: the need for an experimental approach

The term 'safety voice' is used as a broad label to, confusingly, encompass a behaviour and its counterpart: safety voice (i.e., raising safety concerns) and safety silence (Manapragada & Bruk-Lee, 2016; Morrow et al., 2016; Okuyama et al., 2014; Van Dyne et al., 2003). 'Safety voice' often relates to raising safety concerns, which is the act of speaking-up about safety issues, through informal or formal communication channels, to a variety of targets (e.g., management, co-workers, the public), with the intention to mitigate harm from a situation perceived to be dangerous (Tucker et al., 2008). Through doing this, people communicate safety issues with the aim of creating a shared perception of the risk and, ultimately, avoiding the danger (Okuyama et al., 2014). Safety silence, the 'non-voicing' type of safety voice, is defined as the active withholding of safety concerns (e.g., Okuyama et al., 2014), and is thus different from the simple absence of speaking-up: this can follow from not having safety concerns (i.e., 'unconcerned silence').

The concept of safety voice emerged from the literature on employee voice and silence (Morrison, 2011, 2014; Van Dyne et al., 2003), and appears similar. Yet, voice behaviours (in the broadest sense) can be distinguished based on message content (e.g., Liang et al., 2012;

Morrison, 2011), and the safety voice literature is characterised by a narrower concern (i.e., limited to prohibiting harm from safety issues), broader application (i.e., beyond organisational environments), more severe outcomes (e.g., fatalities), and has established different antecedents across levels of analysis (e.g., expected impact of harm, safety knowledge, workload, national culture; Noort et al., 2019b). The message content of safety voice relates to the avoidance of harm based on perceived risks, and arguably types of harm may be distinguished: the prevention of physical (e.g., injuries, accidents), psychological (e.g., bullying, harassment), social (e.g., ostracism, unpleasant interactions) or ethical harm (e.g., loss of autonomy; Marshall, 1996). These issues are important to safety voice researchers and practitioners as they can contribute to unsafe outcomes (e.g., bullying can create a poor safety culture), yet physical harm may be easiest to operationalise (i.e., it is closer to a hazard, less ambiguous, easiest to manipulate), and other types of harm may occur beyond (potential) hazards.

Researching safety voice for academic or practice-based purposes is complex due to the elusive and sensitive nature of the phenomenon. Safety voice is a spontaneous response to hazards occurring in natural environments (e.g., wobbly stepladders, incorrect aircraft air pressure settings), and systematic behavioural observations can provide valuable insights into the dynamic social and physical context in which people raise safety concerns (Mulhall, 2003; Reiss, 1971; Rydenfält et al., 2015; van Schagen & Sagberg, 2012), real-time patterns of behaviour (e.g., attention; Lappi et al., 2017; Waller & Kaplan, 2016), demographic variations (Pérez-Tejera et al., 2018) or how people feel and act when they speak-up without having to rely on post-hoc reports (e.g., Mastrofski et al., 1998; Murphy & Dingwall, 2007), and may reveal stronger effects (Brodin et al., 2016). Yet, within natural environments, it is difficult to i) observe short-lived and spontaneous behaviours that may not occur frequently (Mastrofski et al., 2010) in a resource efficient way (i.e., many resources are needed to capture brief moments of speaking-up/remaining silent; Reiss, 1971), ii) record behaviours in a standardised way (e.g., across unsafe situations), iii) assess the riskiness of a situation and whether people are withholding a safety concern (or did not understand the gravity of the situation), or iv) ensure participants are not changing their natural behaviour (Nichols & Maner, 2008). A notable exception to these limitations of naturalistic observations are cockpit voice recordings, but to-date they have received limited empirical study in terms of safety voice (cf. U. Fischer & Orasanu, 2000).

To overcome the challenges of observing safety voice, practice-based investigations (e.g., inquiries, accident investigations; Francis, 2013, 2015; Rogers, 1986) and the vast majority of academic investigations into safety voice (i.e., a systematic review indicated 76% of academic publications; Noort et al., 2019b) utilise methodologies that obtain data from participant reports on whether they or their supervisees raised or withheld safety concerns. For example, through participants providing statements during inquiries (e.g., Francis, 2015), stating their imagined response to a vignette scenario (Schwappach & Gehring, 2014c), recalling scenarios in which they held a safety concern and communicated this to others (e.g., Schwappach & Gehring, 2014d), or completing survey scales that elicit agreement with statements about imagined or generic scenarios (Delisle et al., 2016; Gkorezis et al., 2016). Applications of these methodologies for academic and non-academic purposes have enabled the identification of lay rationales for safety voice, contributing factors to major incidents, cross-sectional comparisons (e.g., across organisational departments), and testing of interventions to alter lay perceptions of the likelihood, with practitioners supplementing academic conclusions through providing better access to people involved in incidents, subject-matter expertise and faster publication of lessons learned (when conclusions are published).

Table 4.1. Methodological shortfalls, needs, and experimental solutions for the investigation of safety voice.

Shortfall	Need	Experimental solution
Behavioural nature of safety voice		
1. Reports provide no behavioural data	Reproduce safety voice behaviours	Create a situation to elicit speaking-up and silence
2. Few methods have operationalised safety voice as emerging from a clear hazard and safety concern	Operationalise a hazard that elicits a single safety concerns and behavioural response	Present a single hazard and ascertain safety concerns
3. Participants cannot be exposed to real hazards and cannot be aware their decisions on safety voice are observed	Minimise potential harm to participants while they believe risk is real	Manipulate the perception of risk, not real risk, using deception procedures
The reliance on memory and imagination		
4. Reports provide inaccurate data on safety voice	Operationalise measures that directly observe safety voice behaviours	Record behaviours through observation (in-person/recording)
5. Floor and ceiling effects can bias estimates of behaviour	Provide measures that enable sufficient variance and observe speaking-up and silence	Calibrate safety concerns to elicit speaking-up and safety silence
The relationship with other variables		
6. Reports provide limited insights into causal relationship between safety voice and antecedents and outcomes	Provide methodologies that can establish and replicate causal relationships	Build a protocol that can manipulate variables of interest
7. Reports on safety voice may be subject to structural confounds introduced through sampling	Minimise the influence of unintended contextual confounds	Sample participants using random procedures
8. A third outcome variable is created if alternative mitigations are possible	Establish a method that limits alternative hazard mitigations to speaking-up and silence	Minimise alternative mitigations of the hazard
9. Relationships may not be reliable over time	Protocols need to enable direct replication and falsification	Provide a clearly specified study protocol

However, there are limitations in the use of report-based methodologies to investigate safety voice. Reports have limited applicability for addressing situational factors (e.g., personal relevance of risk, group dynamics, previous history of raising safety concerns) and mechanisms (e.g., decision-making on risk) that can shape safety voice, and perhaps paradoxically, request people to speak-up about whether they remained silent. Reports on safety voice are always at least one-step removed from the actual behaviour of raising or withholding safety concerns, are over-reliant on imagining or recalling behaviours, and cannot provide predictive insight into how safety voice relates to antecedents and outcomes. Accordingly, the validity of the research remains uncertain, and alternative methodologies focussing on actual behaviour are required to validate findings, and evidence interventions. The use of experiments in related domains (e.g., bystander intervention; P. Fischer et al., 2011) suggest these methods can provide a way to overcome the unique challenges of studying safety voice in hazardous situations. We thus propose that the shortfalls of safety voice methods (summarised in Table 4.1) can be overcome through the development of an experimental methodology that: i) captures the behavioural nature of safety voice; ii) avoids the reliance on memory and imagination; and iii) explores the relationship to other variables as potential causes.

The behavioural nature of safety voice

Research on safety voice has emerged due to recognition that, in high-risk situations, raising concerns is critical to avoiding accidents. Case study investigations have revealed acts of raising and withholding of safety concerns as critical determinants of harm in dangerous situations (e.g., Cocklin, 2004; Moorhead et al., 1991), and the phenomenon is highly behavioural. It typically involves an individual (e.g., an employee, patient, concerned stakeholder) having a concern about a safety issue, and then raising it with another party (e.g., supervisor, doctor, colleagues) in order to prevent harm, or holding back from raising the concern altogether (silence). Yet, and despite the recognised importance of raising safety concerns for avoiding accidents (and silence in allowing accidents; Francis, 2013; Moorhead et al., 1991; Reader & O’Connor, 2014; Tarnow, 1999), investigations into this phenomenon have frequently assumed that reports correspond to real-world behaviour, and are subject to the same mechanisms that drive safety voice (Del Boca & Noll, 2000).

This is problematic because of: i) the often-observed gaps between reports and actual behaviour (e.g., Sheeran, 2002); ii) the lack of behavioural data upon which to base findings and interventions (Weathington et al., 2010); and iii) the low fidelity of actions and context (i.e., operationalisations do not correspond to the behaviour and risky environment; Stoffregen et al., 2003). Accordingly, it remains unclear to what extent safety voice behaviours differ from report-based data and should be observed directly, in a standardised way (i.e., reports may not acquire behavioural data; Shortfall 1), or conceptualised, operationalised and measured as emerging from clear hazards that cause safety concerns (i.e., safety concerns have not been measured alongside safety voice behaviours; Shortfall 2). Establishing this is important for generating accurate baseline data on safety voice (e.g., the average rates of people that are concerned about a hazard and speak-up or remain silent), clarifying the relationship between presented hazards and the extend that these cause concerns, and for generalising and predicting safety voice behaviours.

However, to date, 76% of the safety voice literature (Noort et al., 2019b) has focussed on willingness to raise safety concerns in general (e.g., agreement to generic questionnaire items), post-intervention changes in safety voice, or the extent of safety voice in response to presented hazards without measuring safety concerns (yet for a safety concern item, see: Schwappach & Gehring, 2014c). For example, high-fidelity training simulations (Hanson, 2017) have specified safety voice as a trainable behaviour, whilst only measuring changes in safety voice in pre- and post-training questionnaires, and studies that have exposed participants to (perceived) hazards such as a senior person engaging in unsafe acts (e.g., Aubin & King, 2015; Barzallo Salazar et al., 2014) or medical emergencies (Reime et al., 2016) have assumed that such hazards should trigger safety concerns (yet do not measure this). Furthermore, where observational data on safety voice has been obtained, measurements have included safety voice into higher level codes (Hughes et al., 2014; Reime et al., 2016), focussed on a tendency to speak-up or remain silent without measuring safety concerns (Kolbe et al., 2012, 2014), assumed knowledge about hazards or their presentation elicited safety concerns (Barzallo Salazar et al., 2014), or presented multiple hazards at once (Hodges, 2018). To our knowledge, no studies have investigated the relationship between observed levels of safety voice and reported safety voice, or to measured safety concerns.[18]

This is important, because safety voice is a highly contextualized behaviour: it is assumed to occur in response to the perception of safety being threatened within a particular context (e.g., cockpit, operating theatre, production line) that can be highly ambiguous (e.g., contrasting information, multiple hazards) and complex (e.g., March & Olsen, 1975). Without collecting data on perceptions of risk within a given context one cannot i) compare across hazardous situations (e.g., threats to patient and aviation safety; Tamuz & Thomas, 2006) and ii) make assumptions about why someone may have remained silent (i.e., unconcerned silence versus withholding of safety concerns due to fear of reprisals), or iv) ascertain whether voice occurred due to concern or precaution (i.e., unconcerned voice). Whilst self-report studies can provide insight on general tendencies for safety voice, insights on how safety concerns elicit safety voice behaviours remain minimal (cf. Schwappach & Gehring, 2014c), and behavioural studies have not measured the risk perceptions of the participants being observed.

To study similar phenomena in other fields, experimenters have designed standardised situations for eliciting participant behaviour: for example bystander interventions (for a meta-analysis see: P. Fischer et al., 2011) or defiance/resistance to authority (Kaposi, 2017; Milgram, 1963; Miller et al., 1995). Within the field of voice more generally, experiments have been

[18] High-fidelity simulations have contributed important insights on the behavioural nature of safety voice, and the effects of training and debriefing (Friedman et al., 2015; Pian-Smith et al., 2009; Weiss et al., 2014, 2018), and indicated that leaders respond less favourably to explicit ways of speaking-up (Krenz et al., 2019) or can delay the onset of nursing speaking-up (Krenz et al., 2020). However, and despite the valuable insights drawn from this work, these studies did not present data on whether scenarios elicited safety concerns (i.e., thus limiting conclusions on safety silence).

used to investigate employee voice for volunteering non-safety related information (Morrison et al., 2015). What is common to these studies is that they create a high-fidelity illusion of an emerging problem that requires a behavioural response (e.g., helping a person falling victim to verbal abuse in a bystander scenario; P. Fischer et al., 2006) without endangering participants. Their benefit is that they allow for a behavioural phenomenon to be investigated in a highly controlled environment, with observations then being contextualised to specific scenarios.

To investigate safety voice, a similar approach would be beneficial, with participants engaging in standardised situations that create a safety concern that can be addressed through speaking-up. This is challenging because participants cannot be exposed to genuine physical harm and, to avoid observer effects and study naturalised behaviour (Nichols & Maner, 2008), participants should not be aware that their decisions on safety voice are being observed (Shortfall 3). These issues can only be addressed through designing scenarios that manipulate *perceived* levels of safety (i.e., hazards that elicit a concern, and a need to intervene), not actual risks, while measuring safety concerns and ensuring participant remain naïve to study goals through deception procedures (Weathington et al., 2010). In particular, designing plausible cover stories is important: in the absence of these invalid data may emerge because participants i) deduce the hazard is fabricated; or ii) believe (correctly) that researchers would need to comply with ethical standards that would prevent the scenario.

In summary, an experimental paradigm is required to investigate safety voice in a controlled, standardised, and generalisable way. A key property of any such paradigm is that it elicits observable safety voice behaviours (i.e., both raising and withholding concerns) through manipulating perceived risk and ascertaining safety concerns (i.e., as opposed to exposure to real physical harm), with deception procedures ensuring that participants are naïve to study intentions, and thus their behaviour is natural.

The reliance on memory and imagination

Insight on safety voice has largely been generated through recalled or imagined (in)action during hazardous instances. Whilst practice-based inquiries have investigated actual incidents (Francis, 2013; Rogers, 1986), typically, it is assumed that participants are accurate in remembering and generalising past behaviours (e.g., Schwappach & Gehring, 2014d), or can imagine how they would respond in a safety-related situation (Schwappach & Gehring, 2014c). These data are then used to explain the factors that influence safety voice (e.g., Nembhard, Labao, et al., 2015), to describe its occurrence (Tucker et al., 2008), and predict future outcomes (Blanco et al., 2009). Yet, the validity of this approach is not self-evident, and correlations often low (Reiss, 1971), with participants in report-based studies having been long-shown as unable, or unwilling, to provide accurate data (Bartlett, 1932; Podsakoff & Organ, 1986). Memories are influenced by a limited ability to recall situations: behaviour can be activated by causes outside of conscious awareness at the time of the behaviour such as scents, posters or semantic primes (Aarts et al., 2008; Custers & Aarts, 2010). Furthermore, distances in time, space, or person (e.g., CEOs reporting whether staff in remote locations raised safety concerns for a system introduced the previous year) can further erode data accuracy, and recalling and imagining behaviours is subject to subject-matter expertise and cognitive biases (e.g., availability heuristic; Schwarz et al., 1991). That is, participants may lack knowledge on what constitutes speaking-up, or be unwilling to accurately report safety silence: reports are constructed based on individual attitudes and perceived social norms regarding safety voice (Bartlett, 1932); individuals may experience dissonance between their ideal self-image as able to speak-up and admitting to safety silence (Baumeister, 1982); and social desirability biases half of survey and interview findings (van de Mortel, 2008). For example, desires to appear a good and ethical employee (or effective manager), may bias participants towards reporting speaking-up over safety silence, especially when harmful outcomes occurred in serious or obvious safety situations.

Moreover, recalling and imagining safety voice provides limited scope for exploring the dynamic context in which it occurs. Safety voice surveys, interviews, and vignettes typically aim to increase realism through recall of previously experienced hazardous scenarios or the presentation of scenarios validated by subject matter experts. However, these scenarios remain limited because i) the hazard environments they present will usually differ in some way from reality (i.e., which hinders recall; Lars & Wolf, 2009); ii) the dynamic of a situation is not present (e.g., task pressures on the participant); and iii) there are no immediate consequences for participants, or safety, at the time of data collection. This means, as shown in other research paradigms (Blass, 1999; Milgram, 1974), a gap may exist between reports and behaviour, and addressing this is important for establishing the triggers of safety voice (e.g. hazard perception), and the contextual factors (e.g. interactions between people and situations) that determine voice: or, indeed, silence.

The above factors potentially erode the accuracy of safety voice data collected through report-based methods (Shortfall 4), which undermines the validity of conclusions assumed from data (Bagozzi et al., 1991), and more specifically, how safety voice is assumed to be operationalized in risky situations. Addressing this is important for establishing the triggers of safety voice, and the contextual factors (e.g., interactions between people and situations) that determine the behaviour. An experimental paradigm focussed on eliciting safety voice can address this limitation through facilitating observations of safety voice (e.g., at the time of data collection, or through video), ensuring these are reliably assessed (e.g., using inter-coder reliability for the extent to which an individual raised safety concerns), with participant post-hoc reports being matched to behavioural data. To achieve this, and undertake meaningful statistical analyses, safety voice experiments need to elicit both safety voice acts (i.e., raising a concern) and silence. Floor (i.e., near-complete silence) and ceiling effects (i.e., near-complete voice) can bias estimates of the behaviour (Shortfall 5), with information about change (e.g., through interventions) being lost at the extreme ends of the scale through data censoring (i.e., relevant data falling beyond the scale end-point; Cox & Oakes, 1984). Though statistical procedures are available (McBee, 2010), a successful experimental paradigm should produce sufficient statistical variance and a moderate degree of speaking-up and silence (i.e., a 50-50 split). Thus, an experimental approach enables direct observations of safety voice behaviours, and provides scope for statistical analyses that can evidence higher construct validity.

The relationship with other variables

Data collection using report-based methodologies typically collect data on safety voice and other variables simultaneously (e.g., in the same survey), and using populations that are not randomised. This limits interpretation of the factors that determine or follow safety voice and silence behaviours.

Investigations using reports provide limited insights into causal relationship between safety voice and antecedents and outcomes (Shortfall 6). Yet, to build interventions, safety voice measures need to establish and replicate causal relationships. Antecedents and outcomes have been linked with safety voice and silence, and evidence suggests that interventions can successfully alter reported levels of safety voice. For example, safety silence increases with perceived social risks (e.g., ramifications of speaking-up; Bickhoff et al., 2016), differences in safety knowledge (e.g., Schwappach & Gehring, 2014b), hierarchical power relations (e.g., Seiden et al., 2006), and, conversely, training on why and how to speak-up reduces silence (Delisle et al., 2016; Hanson, 2017; H. L. Johnson & Kimsey, 2012; Kulig & Blanchard, 2016). Yet, such observations tend to be correlational rather than causal in nature. Additionally, controlled manipulations of safety voice antecedents through vignettes (Anicich et al., 2015; Aubin & King, 2015; Schwappach & Gehring, 2014c) or interventions (Habyarimana & Jack, 2011; Hanson, 2017) are scarce and tend to rely on indirect data rather than behavioural observations.

Furthermore, reports on safety voice may be subject to structural confounds (i.e., variables that are not of interest but covary with independent variables and provide alternative explanations of results; Goodwin, 2008) that may emerge from contextual variables that are introduced through sampling (Shortfall 7; e.g., junior doctors needing longer to accrue subject-matter expertise in part of the included research contexts). To establish valid conclusions, measures need to minimise the influence of confounds and minimise alternative explanations of relationships between antecedents and safety voice and silence. Yet, report-based methodologies have sampled within similar populations (e.g., oncology departments, medical students; Delisle et al., 2016; Schwappach & Gehring, 2014a), and across different populations (Manapragada & Bruk-Lee, 2016), and both sampling practices can be problematic because unmeasured and uncontrolled characteristics of contexts (e.g., workload; Nembhard, Labao, et al., 2015) can provide alternative explanations of patterns in safety voice. Addressing this is important, and a need exists to minimise the influence of unwanted contextual confounds through applying random sampling procedures.

Hence, a need remains to establish methodologies that can address the relationships between safety voice and other variables. The optimal way to achieve this is through safety voice experiments. These can manipulate antecedents (i.e., enabling causal conclusions), randomise participants (i.e., randomising confounds across the groups to eliminate structural influences), and limit participants' influence on hazard mitigation to a choice on whether to speak-up. Critical to an experimental paradigm is that participants should not be able to mitigate physical harm through other means than speaking-up: a third outcome variable is created when alternative mitigations are possible (Shortfall 8). This means that, when participants have a safety concern, safety silence can be determined through absence of safety voice. The field experiment by Barzallo Salazar and colleagues (2014) showed how surgeon communication style predicts medical students' tendency to speak-up, yet the field experiment did not assesses safety concerns and thus cannot distinguish concerned and unconcerned silence, and because relationships between psychological variables may not be reliable over time (Shortfall 9; Gergen, 1973) a need remains for available experimental protocols that enable the direct replication and falsification of findings (Earp & Trafimow, 2015) in laboratory settings.

The current article

We propose the first experimental paradigm for investigating safety voice in laboratory environments, and establish and evaluate it across three studies in order ensure the protocol meets the nine requirements reported in Table 4.1 that address the shortfalls of current safety voice methodologies. Through doing this, we aim to advance safety voice research by i) enabling a behavioural approach, ii) moving away from a reliance on recall and imagination, and iii) supporting the investigation of causal relationships between safety voice and other variables, which can be used as a basis for intervention.

Below, we describe the 'Walking the plank' paradigm that we have developed for investigating safety voice. We then report on the three studies used to refine and iterate the paradigm, alongside the observations about safety voice yielded from these studies.

The 'Walking the plank' paradigm

Our proposed paradigm for assessing safety voice, the 'Walking the plank paradigm' introduces a decision-point for participants in which they are faced with a hazard (a plank with the potential to break when walked on), and need to decide to either raise their safety concern (and experience any consequences of safety voice) or remain silent and let the situation run its course (with potential harmful implications for victims of the hazard). The paradigm's title is a reference to the naval practice of coercing victims to walk off a plank, plunging into the open sea and certain doom. The parallel is in the fact that perpetrators felt abdicated of responsibility because the victim ostensibly killed themselves (i.e., for onlookers, it was an act of safety silence rather than murder). Our Walking the plank paradigm is generic, and its realistic perceived consequences and randomisation of participants provide for a confound-free assessment of safety voice that enables generalisable conclusions. Before settling on a viable scenario, we considered and abandoned four hazardous scenarios for the experimental investigation of safety voice: crossing a busy road (i.e., the real risk was considerable), faking a terrorist threat (i.e., too politically sensitive; likely to upset participants), interacting with loose electric wiring (i.e., the hazard could be mitigated by the participant through alternative means than safety voice such as unplugging the equipment), and ordering participants to provide approval for future hazardous experiments (i.e., difficult to ascertain risk perceptions; no immediate consequences at time of data collection).

The final scenario involved a person walking across a plank with a perceived low weight limit in the context of an alleged creativity task (the cover story). We chose this hazard because we could manipulate the perception that the plank might break (by having a bendy plank and stating a weight limit) while using a plank that was actually safe. Furthermore, it enables experimental control of variables of interest (e.g., self or other walking on the plank), safety knowledge (i.e., provided information regarding the maximum load of the plank), a plausible cover story (i.e., participation in a creativity task to evaluate and test creative uses of a wooden materials), evaluative mindsets (i.e., participants evaluated aspects of the task), standardisation of the hazard (i.e., consistent materials and research assistants), testing of risk perceptions and safety concerns (i.e., perceived maximum load of the plank and the person sitting/walking on it), a straightforward and resource efficient replication by others, and a systematic observation of the linguistic nature of safety voice (this is beyond the scope of the current article). In this article we show that this paradigm meets our nine criteria.

To test the scenario, we iterated it across three studies. Our goal was for the paradigm to meet the nine requirements (see Tables 3.1 and 3.2) of an effective safety voice experiment. Demonstrating and reporting on this process is important for i) enabling the effective application of the Walking the plank paradigm (e.g., it highlights potential challenges for future research), ii) supporting open science (i.e., protocol histories enable more direct replication; it acknowledges safety voice experiments are challenging and that the final version emerged from addressing this) and iii) supporting future research on safety voice (i.e., it illustrates how amendments to the paradigm can be made and evaluated).

Through the course of three studies (their characteristics are summarised in Table 4.2), we illustrate that the Walking the plank paradigm meets the requirements for safety voice and silence experiments. In brief, in study 1 we demonstrate that the paradigm can elicit safety voice behaviours in a safe, controlled and randomised laboratory environment. In study 2 we refine the protocol and demonstrate it is possible to elicit safety silence. In study 3 we further refine the protocol to enable sufficient risk perceptions and explore the nature of safety voice behaviours.

Table 4.2. Protocol characteristics of study 1, 2 and 3.

Protocol characteristic	Study 1	Study 2	Study 3
Hazard			
Hazard	Sitting	Sitting	Walking
Plank material	Pinewood	Plywood	Plywood
Stated maximum load of plank	45kg	42kg	30kg
Presence of broken plank	Yes	Yes	No
Implicit risk condition included	Yes	No	No
Victim of hazard	Condition (Participant/RA: ns)	RA	RA
Ideas evaluated by participant	Seesaw, Shelving, Door, Juggling, Chair/bench, Slide	Shelving, Mirror, Juggling, Bench, Piece of art	Shelving, Mirror, Juggling, Footbridge, Piece of art
Risk perception calculated	Yes (participants' weight)	Yes (estimated RA's weight)	Yes (estimated RA's weight)
Reported safety concerns in wrap-up questionnaire	No	Yes	Yes
Safety voice			
Direct observation of safety voice behaviours	Yes	Yes	Yes
Observation of safety silence behaviours	No	Yes	Yes
Reported safety voice in wrap-up questionnaire	No	Yes	Yes
Research Assistant (RA) actions			
RA avoids hazard upon safety voice	Not manipulated	Yes	Yes
RA perceived to be naïve to maximum load	Not manipulated	Yes	Condition (yes/no: ns)
RA indicates to respond negatively to speaking-up	Neutral	Condition (yes/no: ns)	Neutral
Perspective taking with RA	Not manipulated	Not manipulated	Condition (be objective/ imagine self as other: ns)
Questionnaires			
Wrap-up questionnaire	Yes	Yes	Yes
Demographic questionnaire (submitted before study)	Yes	Yes	Yes

ns = main-effect of the manipulation was not significant.

Study 1

The aim of study 1 was to establish the protocol for the Walking the plank paradigm (initially 'sitting on the plank'), and provide a first evaluation. Within the guise of a creativity task, participants experienced a perceived hazard designed to elicit safety voice behaviours (i.e., being asked to sit on a plank with a risk of breaking under heavy load). The goals of study 1 were to i) test whether the paradigm could sufficiently elicit safety voice behaviours in response to potential physical harm from breaking the plank; ii) present a perceived, not actual hazard; iii) observe safety voice directly; iv) apply participant randomisation and deception procedures; and v) introduce the experimental manipulation of variables (i.e., minimising harm, hazard presentation, hazard awareness, deception, victim identity) for determining safety voice.

Method

Protocol

A 2(safety: unsafe-control) * 2(victim: participant-research assistant) design was employed. Participants were invited to a study about 'creativity' and allocated to study conditions using double blind and random procedures. The study consisted of three stages. First, participants completed a 5-minute 'creativity task' in which they had to design creative uses of a pinewood plank (L: 120cm, W: 20cm, H:1.8cm) and four blocks of wood. The instruction read: "In this room you find a plank and four pieces of wood. In the box below, write down how you could use a plank and four pieces of wood. Try to be creative and think of as many solutions as you can. You have five minutes." Second, in an interaction with a research assistant, the participants were instructed to undertake and rate the feasibility and creativity of each idea, but were informed that they would test the previous participant's ideas (a standard set: seesaw, shelving, door, juggling, chair/bench, slide) which included a hazardous idea (i.e., 'chair/bench'). Upon re-entering the room, the research assistant stated: "The next stage involves testing these ideas for two things: feasibility and creativity. However, your ideas will be tested by the next participant, and now the ideas of the previous participant are tested". Finally, participants completed an electronic questionnaire (including manipulation checks for hazard awareness and naivety to study hypotheses, and unpresented exploratory variables), after which they received a full debrief.

To present the hazard, and elicit a behavioural response, the instruction for the creativity task included a note on the maximum load of the plank (i.e., 'Please note: the plank can carry a maximum load of 45kg/99lbs/7.1 stone)'; unsafe condition), or no additional note (control condition). Furthermore, a broken version of the plank in the room reinforced this information. In reality, the plank was able to hold at least 125kg. When testing the previous participant's creative ideas, the participant was prompted by the research assistant to place the plank across two chairs (their location marked discretely on the floor) with a gap for a third chair between them. The research assistant then made clear their intention to test the feasibility of the bench through sitting on it (e.g., "Okay, let me test this") or requested the participant to sit (e.g., "Could you please demonstrate?"). The emphasis of the protocol was to observe any subsequent speaking-up or silence behaviour. The protocol concluded with the participant completing a questionnaire.

Ethical approval was obtained for all studies from LSE's research ethics committee (#000540), and informed consent was required from participants before commencing. To comply with data regulations, anonymous data storage to enable future research was included as a separate question.

Participants

129 participants ($N_{females} = 85$, $N_{students} = 98$) were recruited from a pool including students and the general public. Participants were spread in age ($M_{(sd)} = 26.57_{(7.56)}$) and weight ($M_{(sd)} = 64.81kg_{(14.41)}$). On a 5-point Likert scale (with 1 = low), participants indicated they had no expertise on timber ($M_{(sd)} = 1.67_{(1.03)}$), or whistleblowing legislation ($M_{(sd)} = 1.48_{(.83)}$), and safety voice did not correlate with demographic variables (i.e., student status, gender, age, social economic status, class, education, expertise on timber/whistleblowing, nationality, language). One participant was dropped from analyses because the protocol was not followed.

Measures

Manipulation checks. Perceived risk was calculated from two items in the questionnaire that followed the scenario (i.e., kilograms of participants' own weight minus the estimated plank's maximum load). This measure addressed that the plank's maximum load would not pose a safety issue without a person sitting on it. One participant's estimation of the maximum load of the plank (i.e., 292kg) was removed based on a Cook's test identifying the response as an outlier (i.e., for the effect of the safety condition on risk perception; Cook = .50). The questionnaire asked whether participants noticed anything odd during the study.

Safety voice. A direct observational measure of safety voice was used. Safety voice (1) was coded if the participant questioned whether testing the bench was a good idea and/or alternative action might be more appropriate (e.g., "Did the instruction not state a maximum of 45kg?"; "This would be feasible for a child, not for adults"), before the chair/bench was tested. Otherwise the participant's behaviour was recorded as 'no voice' (0). Through discussing examples, research assistants were trained to recognise whether statements intended to prevent a situation in which someone sat on the plank and might break it. The first author made a final decision through watching video recordings when research assistants were unsure on how to code participants statements.

Prohibitive employee voice. Three items from Liang and colleagues (2012) were adapted to the laboratory environment to explore overlap with safety voice (on 5-point Likert scale, with 5 indicating strong agreement): 'I pointed out problems when they appeared, even if that would hamper relationships with others'; 'I advised others against undesirable behaviours that might hamper the task'; 'I highlighted problems that might cause serious issues'.

Results

Manipulation check. The paradigm's safety manipulation created a perception that sitting on the plank would break it (i.e., weight difference between person sitting and plank's maximum load ≥ 0kg). The perceived maximum load of the plank was 13.96kg lower in the unsafe condition ($M_{(se)}$ = 48.84kg$_{(2.97)}$), $F(1,127) = 4.39$, $p = .04$, $\eta^2 = .03$, *observed*

Table 4.3. Safety voice behaviours for Study 1 (unsafe condition).

	Perceived risk		Perceived no risk		Total	
	N	$\%_{(SE)}$	n	$\%_{(SE)}$	n	$\%_{(SE)}$
Voice	21	$42_{(7.1)}$	4	$33_{(14.2)}$	25	$40_{(6.2)}$
Silence	29	$58_{(7.1)}$	8	$67_{(14.2)}$	37	$60_{(6.2)}$
Total	50	$81_{(5.1)}$	12	$19_{(5.1)}$	62	$100_{(-)}$

Percentages total 100% within a column, except for the total of perceiving (no) risk.

power = .55. The perceived risk for the unsafe condition ($M_{(se)}$ = 19.60kg$_{(3.26)}$) was non-zero, $t(59) = 6.00$, $p < .001$, higher than the control condition ($M_{(se)}$ = -1.32kg$_{(5.97)}$), $F(1,127) = 7.72$, $p = .006$, $\eta^2 = .06$, *observed power* = .79, and led 81% of participants in the unsafe condition (95CI: 71-91%) to think the plank would break, $t(61) = 15.94$, $p < .001$. Illustrating successful deception, no participant guessed the true nature of the study.

Safety Voice. The safety manipulation successfully elicited safety voice. Whilst some participants raised safety concerns in the control condition (i.e., 20% spoke-up; 95CI: 10-29%), $t(66) = 3.99$, $p < .001$, participants were 2.76 times more likely to raise safety concerns against sitting on the bench when information regarding an unsafe maximum load was provided, $Wald(1) = 6.12$, $p = .01$. Yet, and despite the success of the manipulation to create risk perceptions for 81% of participants in the unsafe condition, a considerable proportion of participants in the unsafe condition did not raise a concern (60%; 95CI: 48-73%), and this held when participants without a perceived risk were accounted for: 58% (95CI: 44-72%) remained silent about their perceived risk (see Table 4.3). Furthermore, in the unsafe condition, 33% (95CI: 2-65%) of participants raised a safety concern despite not perceiving a risk, $t(11) = 2.35$, $p = .04$, and perceiving risk was not related to safety voice, $\chi^2(1) = .30$, $p = .58$. However, whilst the safety manipulation caused differences in safety voice, no influence was found on prohibitive employee voice, $F(1,127)$s < 1.29, *p*s > .26, and no correlation existed with observed safety voice, *r*s > |-.10|, *p*s > .25. This suggests that hazards differentiate safety voice but the relationship between risk perception and safety voice is not straightforward. A need thus exists for improved safety concern measures. Finally, the identity of the victim (i.e., participant vs research assistant) did not influence safety voice, *ns*[19].

Discussion

Study 1 demonstrated that the paradigm enables i) the reproduction of safety voice behaviours in response to a hazard (speaking-up only); ii) the presentation of a perceived, not actual, hazard; iii) the direct observations of safety voice; iv) participant randomisation to minimise alternative explanations; and v) experimental control over study variables (i.e., minimising harm, hazard presentation, hazard awareness, deception, victim identity). Furthermore, it suggested that the relationship between risk perceptions and safety voice is not straightforward, and participants can remain silent when perceiving a risk, or speak-up when not perceiving a risk.

However, study 1 did not fully illustrate five requirements for safety voice experiments. First, participants raised safety concerns when demonstrating the seesaw and slide ideas, thus presenting multiple hazards and potentially producing unmeasured spillover effects. Second, it was not clear whether the perception of risk made people concerned about the hazard: it is not self-evident that safety concerns emerge from participants' body weight, or that the application of this weight to a plank with a low capacity always leads to concerns, and in order to demonstrate safety silence (i.e., the withholding of safety concerns) experiments need to establish optimal measures to establish safety concerns. This is important, because, third, whilst safety voice behaviours were observed, these emerged for people with and without perceptions of the plank potentially breaking, and in the absence of clear safety concern measures it is unclear whether a lack of voice meant safety concerns were withheld (i.e., participants might

[19] An online pilot study (n = 88), that asked participants to rate a video of the scenario, suggested that participants perceived the hazard equally unsafe dependent on whether they or someone else would be the victim, $F(88,1) = .03$, $p = .86$, $\eta^2 = .00$.

not have been concerned about harm despite a perceived likelihood of the plank breaking). Fourth, the proportion of safety voice acts was low and could be improved to prevent floor effects. Finally, when participants were victim, they occasionally mitigated the hazard by keeping weight on their feet and thus not fully sitting on the plank (creating a third outcome variable).

Study 2

Study 2 aimed to address the issues raised in study 1 through amending the risk perception measures to enable the observation of safety silence (i.e., calculated based on the person sitting on the bench and triangulated with an item on having a safety concern); eradicating safety voice for multiple hazards; improving the manipulability of the perceived physical risk to elicit stronger responses (i.e., lowering the weight limit; using a bendy plank; creating sufficient variance in safety voice and silence); and minimising alternative ways to mitigate physical harm following from breaking the plank[20].

Methods

Protocol refinements

The protocol in study 1 was followed, albeit with five adjustments. First, the observation of safety silence was enabled through an altered risk perception measure and self-report safety voice questionnaire item to obtain additional data and ascertain whether the scenario led to subjective safety concerns. Second, to increase the perceived risk of physical harm, the maximum load was lowered slightly to 42kg (93lbs, 6.6 stone) and the pinewood plank was replaced by a more bendy plywood plank of the same proportions (still capable to withstand at least 125kg in reality). Third, to eliminate other perceived hazards from the protocol, three

[20] The manipulation of one experimental variable was explored (i.e., high and low expectations of negative consequences of voice), but this is not discussed here because the condition was randomised and produced no significant main-effect.

ideas (i.e., seesaw, door, slide) were replaced with two new ideas (i.e., mirror, piece of art). Fourth, to ensure that the hazard could not be mitigated through not fully sitting on the plank, the research assistant sat on the plank. Finally, based on a pilot study, only the unsafe scenario was included[21].

Participants

Sixty-nine participants were recruited ($N_{females} = 50$; $N_{students} = 62$; Age $M_{(sd)} = 25.52_{(.61)}$). Participants had no expertise on timber ($M_{(sd)} = 1.54_{(.11)}$), or whistleblowing legislation ($M_{(sd)} = 1.62_{(.12)}$), and demographic variables were not associated with safety voice measures.

Measures

Manipulation check. Perceived risk was based on the estimated weight of the research assistant (i.e., estimated weight of the research assistant's above the plank's maximum load). Furthermore, a dichotomous item asked whether participants were concerned regarding the demonstration of the bench (answered as: yes/no). One participant's estimation of the maximum load of the plank (i.e., 200kg) was removed based on a Cook's test identifying the response as an outlier (i.e., Cook = .09).

Safety voice. Safety voice acts were observed as a dichotomous variable, described in study 1. Furthermore, participants' self-reported safety voice was measured as a dichotomous variable (i.e., did you raise a safety concern regarding the demonstration of the bench idea: yes/no). Safety silence was operationalised as participants who said they held a safety concern but were not observed to raise it.

[21] A lab-based pilot study (n = 38) demonstrated that the mere presentation of the message regarding the maximum load of the plank was sufficient to reproduce safety voice behaviours, $OR = 17.00$, $Wald(1) = 6.38$, $p = .01$.

Results

Manipulation check. The safety manipulation created a perception that sitting on the plank could break it (i.e., excess weight ≥ 0kg): the perceived maximum load of the plank (M(se) = 60.28kg(3.20)) was not statistically higher than the weight of the research assistant sitting on the plank (M(se) = 57.69kg(1.26)), $t(67) = -.87$, $p = .39$, 55% of participants (95CI: 43-67%) perceived that the plank could break, $t(66) = 9.02$, $p < .001$, and 42% (95CI: 30-54%) reported feeling concerned, $t(68) = 7.02$, $p < .001$. The new safety concern measure had a stronger relationship to safety voice than perceived risk: whether participants perceived a physical risk was not related to observed safety voice behaviours, $OR = 1.64$, $Wald(1) = .79$, $p = .37$, yet whether participants reported having a safety concern about the act of sitting on the plank related to safety voice, $\chi 2(1) = 4.14$, $p = .04$, and these people were 3.16 times more likely to be observed to raise a safety concern, $Wald(1) = 3.81$, $p = .05$. Furthermore, safety concerns were predicted by the perceived risk, $OR = 3.61$, $Wald(1) = 5.87$, $p = .02$, indicating an indirect relationship between perceived risk and safety voice behaviours through safety concerns.

Safety voice. Replicating study 1, safety voice behaviours were directly observed, but 71% (95CI: 59-82%) of participants did not raise a concern about the research assistant testing the bench, $t(67) = 5.28$, $p < .001$. Yet, strikingly, and demonstrating concerned silence, this held when participants without safety concerns were not included in the analysis: 57% of participants did not raise their safety concern (95CI: 38-76%), $t(27) = 4.50$, $p < .001$. Furthermore, and suggesting the existence of two additional types of safety voice behaviours (i.e., unconcerned voice and silence), 20% (95CI: 7-33%) of participants raised a safety concern despite being unconcerned, $t(39) = 3.12$, $p = .003$ (see Table 4.4).

Table 4.4. Safety voice behaviours for study 2.

	Concerned		Unconcerned		Total	
	n	$\%_{(SE)}$	n	$\%_{(SE)}$	n	$\%_{(SE)}$
Voice	12	$43_{(9.5)}$	8	$20_{(6.4)}$	20	$29_{(5.5)}$
Silence	16	$57_{(9.5)}$	32	$80_{(6.4)}$	48	$71_{(5.5)}$
Total	28	$42_{(6.0)}$	40	$58_{(6.0)}$	68	$100_{(-)}$

Percentages total 100% within a column, except for the total of (un)concerned. (Missing: 1)

Self-reported safety voice. Supporting the need for direct behavioural observations of safety voice behaviours, participants provided poor report of their behaviours. Reported safety voice related to observed safety voice ($r = .47$, $\chi 2(1) = 14.94$, ps < .001), but it only explained a small proportion of the variance ($R^2 = .22$), and 23% of participants (95CI: 13-34%) misreported their behaviour, $t(67) = 14.76$, $p < .001$, with participants 4.5 times more likely to misreport that they raised a concern, $Wald(1) = 6.25$, p = .01.

Discussion

Study 2 successfully addressed the requirements to observe safety silence, ascertain safety concerns for a single hazard, and minimise alternative hazard mitigations. Furthermore, study 2 showed it is important to include safety concern measures in safety voice experiments, indicated that safety voice consists of four behaviours (i.e., concerned voice and silence; unconcerned voice and silence), demonstrated a gap between observed and reported safety voice, and indicated that participants tend to misreport in favour of speaking-up.

However, study 2 was limited in terms of eliciting safety concerns from the majority of participants, with even fewer participants (as a proportion of concerned participant) raising their concern. The reasons for this are unclear, yet consistent with the wider safety voice literature, may reflect either an unwillingness to voice safety concerns, or a perception that the situation does not merit action (unconcerned silence). In particular, only 42% of participants were concerned about the act of sitting on the plank (and of these 57% withheld their concern). This indicates that for the majority of participants the task was not particularly risky, and for those who did perceive it as risky, it may not have been perceived as sufficiently dangerous to warrant intervention. Thus, to increase engagement in safety voice behaviours, and prevent a floor effect, we decided to further increase participants concern.

Study 3

Study 3 refined the paradigm so that it would meet the final requirements for a safety voice paradigm: to increase the number of participants with safety concerns and produce an equal amount of safety voice and silence acts. It replicated the four safety voice behaviours identified in study 2, and refined the protocol through further reducing the stated maximum load of the plank to 30kg and altering the interaction with the plank to walking the plank, rather than sitting[22].

Methods

Protocol refinements

To improve the number of concerned participants, study 3 refined the paradigm's protocol through replacing the previous participant's idea for creating a bench by a footbridge. Instead of sitting on the plank when it is placed across two chairs, the research assistant made clear he/she would be testing the idea by walking over it (and did so in the absence of safety voice)[23]. The final protocol is presented in an online manual as supplemental material (providing detailed pictures, scripts)[24].

Participants

Seventy-five participants were recruited ($N_{females} = 49$; $N_{students} = 69$; Age $M_{(sd)} = 23.09_{(3.87)}$, missing demographic data: 1 person). In reply to dichotomous questions (i.e., are you an expert

---

[22] The manipulation of two experimental variables was explored (i.e., perceived hazard awareness of the research assistant; perspective-taking), but these are not discussed here because the conditions were randomised and produced no significant main-effect.

[23] An online pilot study (n = 57), that asked participants to rate a video of the scenario, revealed that the act of walking the plank was considered unsafe and likely to lead to undesirable and preventable physical harm by 76% of participants, $t(56) = 7.28$, $p < .001$.

[24] This manual is also presented in Appendix C.

on wood/whistleblowing), no participants reported to be whistleblowing experts and only 1 participant reported to be a wood expert. Demographic variables were not associated with safety voice measures.

Measures

Study 3 adopted the manipulations checks (i.e., perceived risk, self-reported safety concern) and safety voice measures (i.e., observed acts, self-reported) described in study 2. Ten additional exploratory items (on a 5-point Likert scale, with 5 indicating strong agreement) were included: "I felt I might be seen as a trouble-maker when I spoke up" (Wei et al., 2015); "I felt obligated to raise any concerns I had" (Liang et al., 2012); "Right now, I worry about making mistakes" (Carver & White, 1994); "I felt I might offend the RA by questioning the way things were done"; "I felt the RA might bring out the worst in me"; "I felt uncomfortable to speak up about concerns I had"; "I had a concern about something that I thought the RA was not aware of"; "I had more information than the RA"; "I withheld my opinions"; "I don't feel very sorry for any problems the research assistant might have" (reverse-coded).

Results

Manipulation check. Altering the safety manipulation to walking the plank improved concerns that the plank would break it (i.e., weight difference ≥ 0kg): the perceived maximum load of the plank ($M_{(se)} = 38.89kg_{(2.93)}$) was significantly lower than the perceived weight of the research assistant walking the plank ($M_{(se)} = 56.76kg_{(1.48)}$), $t(73) = 6.08$, $p < .001$, 82% of participants (95CI: 74-91%) perceived the plank could break, $t(73) = 18.51$, $p < .001$, and 68% (95CI: 57-79%) reported feeling concerned, $t(73) = 12.54$, $p < .001$. This proportion of concerned participants was significantly higher than for Study 2 (i.e., 42% concerned participants), $t(74) = 4.80$, $p < .001$.

Safety voice. The protocol for Study 3 elicited speaking-up for 44% (95CI: 33-56%) of participants, $t(74) = 7.63$, $p < .001$, and, demonstrating safety silence, this held when

Table 4.5. Safety voice behaviours for Study 3.

	Concerned		Unconcerned		Total	
	n	$\%_{(SE)}$	n	$\%_{(SE)}$	n	$\%_{(SE)}$
Voice	26	$51_{(7.1)}$	7	$29_{(9.5)}$	33	$44_{(5.8)}$
Silence	25	$49_{(7.1)}$	17	$71_{(9.5)}$	42	$66_{(5.8)}$
Total	51	$68_{(5.4)}$	24	$32_{(5.4)}$	75	$100_{(-)}$

Percentages total 100% within a column, except for the total of (un)concerned.

unconcerned participants were excluded, $t(50) = 7.21$, $p < .001$: participants with safety concerns were split between those who raised (i.e., 51%) or withheld (i.e., 49%) their safety concerns, and, providing sufficient variation in safety voice and silence, this split was not different from a 50-50% split, $t(50) = .14$, $p = .89$, and this was true for concerned and unconcerned participants, $\chi^2(1) = 3.15$, $p = .08$. Furthermore, and providing further support for the existence of (un)concerned voice and silence, 29% of participants (95CI: 10-49%) raised a safety concern despite being unconcerned, $t(39) = 3.08$, $p = .01$ (see Table 4.5).

A MANOVA (using Pillai's trace) suggested that participants who displayed either concerned voice, concerned silence, unconcerned voice or unconcerned silence responded differently to ten exploratory questionnaire items, $V = .86$, $F(30,192) = 2.56$, $p < .001$, $\eta^2 = .29$, *observed power* = .91, and separate ANOVAs confirmed this, $Fs(3,71) \geq 3.41$, $p\text{s} \leq .02$, $\eta^2 s \geq .13$, *observed power* ≥ .75. Post-hoc analyses suggested that people who raised their concerns were less fearful, more caring and thought they had more information compared to those who withheld their concerns: they were less likely to fear being seen as a trouble-maker, $MD = -.79$, $p = .01$, offend the RA, $MD = -.79$, $p = .05$, or making mistakes, $MD = -.86$, $p = .01$, state to withhold their opinions, $MD = -.92$, $p = .01$, feel sorry for any problem the research assistant had, $MD = -.77$, $p = .02$, or obligated to raise concerns, $MD = -.81$, $p = .02$, and think they had more information than the RA, $MD = -.81$, $p = .03$. Furthermore, and suggesting a lack of safety concerns might be due to feeling less responsible for the research assistant, in comparison to those who raised their concerns, those who spoke-up despite being unconcerned felt less obligated to raise concerns, $MD = -1.48$, $p = .004$, and less sorry for the research assistant's problems, *MD* = -1.16, *p* = .02, and a marginally significant trend suggested they might have less concerns that they think the research assistant was not aware of, *MD* = -.1.24, *p* =.06. Providing further evidence that concerned participants who remained silent were fearful, they were more likely than those who raised concerns despite being unconcerned to state they withheld their opinions, *MD* = .1.27, *p* =.03, and feel uncomfortable to speak-up, *MD* = .1.15, *p* =.04. Finally, and suggesting that making people concerned can improve speaking-up when people display unconcerned silence, those who displayed unconcerned silence were more likely than those displaying unconcerned voice to feel obligated to raise concerns, *MD* = -.1.24, *p* =.04, and less likely than those who raised concerns to perceive a concern that they felt the research assistant was not aware of, *MD* = -.1.03, *p* =.02.

Self-reported safety voice. Replicating Study 2, and supporting the need for direct behavioural observations of safety voice, participants provided poor self-reports of their safety voice behaviours. Self-reported safety voice for Study 3 related stronger to observed safety voice than for Study 2, $r = .64$, $\chi^2(1) = 30.39$, *p*s < .001. However, a considerable portion of the variance remained unexplained ($R^2 = .41$), and 19% of participants (95CI: 10-28%) misreported their behaviour, $t(74) = -4.12$, $p < .001$, but participants' tendency to misreport safety voice acts over safety silence was only a marginal trend, *OR* = 2.78, *Wald*(1) = 2.48, $p = .092$ (bootstrap sample = 1000).

Table 4.6. The illustration of requirements for safety voice experiments across study 1, 2 and 3.

Requirement	Study 1	Study 2	Study 3
Behavioural nature of safety voice			
1. Created a situation to elicit speaking-up and silence	No: unclear whether silence was concerned silence	Yes	Yes
2. Presented a single hazard and ascertained safety concerns	No: poor measure; more than one hazard	No: safety concerns could be increased	Yes
3. Manipulated the perception of risk, not real risk, using deception procedures	No: concerns unclear	Yes	Yes
The reliance on memory and imagination			
4. Recorded behaviours through observation	Yes	Yes	Yes
5. Calibrated safety concerns to elicit speaking-up and safety silence	No: more voice needed	No: still more voice needed	Yes
The relationship with other variables			
6. Built a protocol that can manipulate variables of interest	Yes	Yes	Yes
7. Sampled participants using random procedures	Yes	Yes	Yes
8. Minimised alternative mitigations of the hazard	No: some participants did not fully sit	Yes	Yes
9. Provided a clearly specified study protocol	Yes	Yes	Yes

Discussion

Study 3 successfully addressed the remaining challenges to the paradigm (i.e., creating sufficient safety concerns; producing equal safety voice and silence) through altering the presented hazard from sitting on the plank to walking over it. This amendment increased the number of concerned participants and thus creates ample scope to test interventions for safety voice because the resulting proportion of safety voice (i.e., 50%) could be improved and reduced through the manipulation of safety voice antecedents. Progress of the development of the Walking the plank paradigm across the 3 studies is summarised in Table 4.6.

Furthermore, Study 3 revealed that the four types of safety voice behaviours (i.e., concerned voice, concerned silence, unconcerned voice, unconcerned silence) were associated with different levels of fear, felt obligation and care for the research assistant.

General discussion

Our results establish a novel experimental paradigm for safety voice. Through an iterative process, three studies addressed nine requirements for a valid safety voice experiment (see Table 4.1). The final protocol can facilitate behavioural investigations of safety voice, overcome the reliance on memory and imagination inherent in report methodologies, and allow for the study of relationships between safety voice and other variables. It is also the first generalisable experimental paradigm for safety voice, enables the investigation of (un)concerned voice and silence, can be used to investigate the effect of safety voice interventions, and through focussing on behaviour, can improve the conceptualisation and operationalisation of safety voice.

Conceptualising and operationalising safety voice

Through the process of developing the ‘Walking the plank’ paradigm, insights were drawn on conceptualising and operationalising safety voice, the relationship between having and raising safety concerns, the important role of safety silence, and existence of unconcerned voice and silence.

First, studies of safety voice would benefit from operationalising the phenomenon as observable behaviours in response to safety concerns rather than reportable acts. In our studies, it was notable that the presentation of the hazardous footbridge elicited observable safety voice behaviours, and that these often differed from reported safety voice (i.e., about 1 in 5 participants misreported their behaviour, and participants tended to favour misreporting safety voice over silence). This finding reinforces the problems we raised with report methodologies at the outset, and has implications for conclusions from practice-based and academic investigations. Whilst practice-based investigations occur in response to real hazards that elicited safety concerns, these frequently rely on reports of incidents occurred in the past (e.g., interviews, focus groups; Francis, 2013). Furthermore, existing academic studies using behavioural observations (Aubin & King, 2015; Barzallo Salazar et al., 2014; Hu et al., 2015; Hughes et al., 2014; Kolbe et al., 2012, 2014; Reime, Johnsgaard, Kvam, Aarflot, Breivik, Engeberg, Brattebø, et al., 2016; Sundqvist & Carlsson, 2014) have not addressed the extent that safety concerns lead to speaking-up (e.g., they assumed that an unsafe procedure causes concerns). Other scholars have merely discussed the behaviourial nature (Hofmann et al., 2003; C. Jones & Durbridge, 2016), sampled experiences close to the behaviour (Kines et al., 2010) or provided high fidelity simulation training on the behaviour (Hanson, 2017) without measuring it through behavioural observations. Accordingly, more measures of safety voice should operationalise (or triangulate with) direct observations of behaviours. We showed these behaviours emerged in response to a hazard (i.e., a hazard should be presented, also see below) to prevent physical harm (i.e., safety voice occurs before the hazardous scenario has finished: voice after the scenario may aim to clarify a mismatch between safety information and the results of the scenario).

Second, safety voice is thus rooted in hazard perception: it primarily occurs in response to being presented with a hazardous situation. Yet, the relationship between safety voice and hazards is not straightforward, with hazards emerging through interactions between behaviour and physical contexts, and safety voice occurring when hazard perceptions trigger a safety concern. That is, crucially, whilst the plank of wood became unsafe when the context of its utilisation changed (e.g., from being held up as a mirror, to being used as a low-weight carrying footbridge), our data illustrated that perceiving this as risky only related to safety voice because the risk of breaking concerned participants. Furthermore, we illustrated that i) an objectively safe plank could be perceived as unsafe before someone walked the footbridge, ii) not everyone held a safety concern despite safety information regarding the plank's maximum load and perceptions that walking the plank would break it, and iii) some participants spoke-up about safety without reporting feeling concerned. This indicates that safety concerns are based in risk perceptions that are future orientated (i.e., consequences have not yet occurred), uncertain (i.e., there is no direct evidence available to participants regarding consequences of hazards), and subjective (i.e., desirability of physical harm is a personal preference). Thus, safety voice behaviours are not just indicative of the willingness of a participant to raise a safety concern, but reflect more nuanced and subjective judgements rooted in the uncertainty of perceiving hazards, the interplays between actions and objects, and attribution of desirability.

Third, we learned that safety voice behaviours could be codable as a dichotomous variable (i.e., safety voice was observed: yes/no) but were expressed in different utterances which might be better captured in future as a more continuous variable. For example, people raised their safety concern through stating the facts (e.g., 'there was a 30kg weight limit'), exclaiming concern (e.g., "No, don't do that!"), asking for additional information (e.g. "wait, how heavy are you?"), polite statement (e.g., 'this should be safe for a kid'), and some participants persisted or physically blocked the research assistant from engaging with the plank. Coding these behaviours in a binary manner enabled statistical analyses (e.g., logistic regressions) and conclusions on the extent that participants raised their safety concerns, yet this variety underscores the need expressed by others to understand voice concepts as conversational acts that can be expressed in different ways (Bashshur & Oc, 2014; A. Jones & Kelly, 2014; Kulig & Blanchard, 2016; Lyndon, 2008; Manning, 2006). Binary approaches may oversimply otherwise meaningful utterance as silence (A. Jones & Kelly, 2014), or obscure moderators and drivers of outcomes (Bashshur & Oc, 2014), and crew resource trainings may benefit from understanding the breadth of conversational techniques employed to raise concerns. The Walking the plank paradigm provides a new methodology for exploring this, and particularly enables data collection on variation in utterances to standardised and controlled hazards.

Fourth, we learned that operationalising safety silence behaviours is deeply challenging. Developing strong safety concern measures is important for assessing whether participants

Table 4.7. Safety voice typology.

	Safety Concern	**No Safety Concern**
Voice	Concerned Voice	Unconcerned Voice
Silence	Concerned Silence	Unconcerned Silence

withhold concerns, but ascertaining safety concerns is more difficult than observing safety voice behaviours because concerns are intrinsically subjective (as per the above), and cannot be observed directly. Crucially, and echoing the literature on employee voice (Morrison, 2014), for investigating safety voice it is important to ascertain whether participants were concerned. We addressed this through triangulating risk perception and safety concern measures, and alternative operationalisations of safety concerns may be developed for the assessment of safety silence during or before the exposure to the hazard (e.g., physiological measures).

Fifth, through obtaining data on safety concerns we indicated voice and silence for people who were concerned and unconcerned, and the existence of these four types of safety voice merits investigation and conceptualisation. Unconcerned voice may be explained as verbalised sense-making on safety, caution, or a misrepresentation of being concerned, and unconcerned silence may be a misrepresentation or an unawareness of the hazard. In particular, our results suggested that i) people who withheld their concerns worried more (e.g., about being seen as trouble-maker), cared less about the research assistant, and felt less obligated to raise concerns; ii) those who spoke-up despite perceiving no safety issues felt less responsible for the research assistant; and iii) those who were silent and unconcerned simply did not perceive an issue, but would have felt responsible for raising it. Thus, our paradigm enables the development of safety voice into a two-dimensional typology for (un)concerned voice and silence that is rooted in social interaction and sense-making (see Table 4.7), and this may resemble signal detection typologies (i.e., hit, miss, false positive, false negative; Nesse, 2005).

New directions for the investigation of safety voice

The literature of safety voice has established a considerable body of findings, and the 'Walking the plank' paradigm enables four potential research questions.

Can safety voice antecedents be replicated?

The paradigm can be used to replicate empirical findings within an experiment setting. We illustrated that safety voice behaviours can be elicited and directly observed, and that self-reports of safety voice were imperfect and biased towards speaking-up. This might raise doubts on report-based evidence regarding the relationship between antecedents and safety voice, and interventions based on these conclusions. For example, self-report and correlation studies indicate that expectations of negative consequences of speaking-up (e.g., Bickhoff et al., 2016) and power hierarchies (e.g., Seiden et al., 2006) can lead to withholding safety concerns. However, whilst power hierarchies have been experimentally manipulated (Anicich et al., 2015; Barzallo Salazar et al., 2014; Schwappach & Gehring, 2014c), simulated (Aubin & King, 2015; Reime, Johnsgaard, Kvam, Aarflot, Breivik, Engeberg, Brattebø, et al., 2016), trained on (Hanson, 2017), or discussed as part of single-cases (Liao et al., 2014), to date causality remains unclear because no studies have simultaneously i) manipulated power hierarchies (e.g., through creating a control condition) and ii) obtained behavioural data on safety voice (treating reports of the behaviour as empirically different from the behaviour itself). Using the 'Walking the plank' paradigm, research can establish whether current findings are upheld in an experimental setting, and establish the causal relationship between safety voice and other variables.

What characterises safety voice behaviours?

We demonstrated safety voice as a behavioural phenomenon, and this opens up new questions for how the behaviour can be characterised. Our experimental paradigm enables the identification of the nature of safety voice behaviours, and especially opens up the investigation of question regarding sense-making, decision-making, physiological mechanisms and linguistic expressions. In particular, although the literature has conceptualised safety voice (i.e., raising concerns held) and silence (i.e., withholding concerns held), we provided the first demonstration and initial conceptualisation of the act of raising safety concerns (and staying silent) when individuals are unconcerned. Through refining the conceptualisation of (un)concerned voice and silence, and testing its predictive power in relationship to safety voice antecedents or outcomes would provide clarity on the nature of safety voice behaviours, and enable targeted interventions.

In addition, we showed that safety voice appears distinct from employee voice. Speaking-up to change expected outcomes (as the verbal behaviour labelled 'voice') is at the core of employee and safety voice, and scope for integration may exist under conceptual overlap (Wilkinson et al., 2019). However, conceptually employee voice includes a broader set of behaviours (i.e., promotive and prohibitive; Liang et al., 2012) than safety voice (i.e., preventing harm is prohibitive in nature), and we found no empirical support that safety voice is a sub-type of employee voice. This may be because, in hazardous situations, safety voice emphasises the prevention of harm based on an assessment of perceived risks that can be ambiguous (e.g., because they are yet to occur, have not been noticed, it is not clear who is responsible, or people are discouraged to raise concerns). This may prompt sensemaking on potential harms (with or without the interlocuter) and felt responsibilities for harm-prevention, and a clearer responsibility and need for sensemaking may lie in warning others than challenging their task-related choices. This underscores the need to consider the content of the raised message through speaking-up (see also: Morrison, 2011), and this extends to the breadth of harmful issues that are raised through safety voice. Devising studies that manipulate the content of safety voice (i.e., judgements on risk, attributions of desirability, types of harm) provides a way to understand voice behaviours.

Third, the direct observation of safety voice behaviours enables the assessment of the decision-making processes regarding whether to raise concerns or not. Decisions to raise safety concerns may be automatic or deliberative, and the optimal way to make decisions for speaking-up about safety remains unexplored. Evidence suggests that people are more inclined to engage in pro-social behaviours under time-pressure (Rand et al., 2014) or in a state of pro-social disinhibition (van den Bos, Muller, et al., 2011; van den Bos et al., 2009; van den Bos, van Lange, et al., 2011), and intuitive decisions are frequently implicated (Rand & Epstein, 2014) and effective (Kahneman & Klein, 2009) in preventing harm based on recognising patterns in the situation (e.g., fire fighters recognising that smoke patterns indicate a potentially lethal backdraft). Through manipulating the time-pressure, the paradigm may therefore unearth decision-making mechanisms for safety voice.

A fourth area for conceptualising safety voice behaviours is the association of safety voice with physiological measures. Because the proposed paradigm enables the observation of safety voice in-situ, these behaviours can be simultaneously assessed with physiological mechanisms: safety beliefs emerge from embodied experiences (Somerville, 2006) and consequences of silence might manifest physiologically. Scholars may therefore explore the generalisation of safety concerns from physiological mechanisms. Our paradigm lends itself for the inclusion of physiological measures (e.g., skin conductance, heart rate, inhalation, gross movement, vocal amplitude, vocal pitch) and this enables the conceptualisation of the physiological mechanisms underpinning safety voice (e.g., arousal) that can be triangulated to safety concerns measures.

Finally, researchers may utilise the safety voice paradigm to examine the linguistic nature of safety voice. We, in line with others in the literature (e.g., Okuyama et al., 2014), treated safety voice as a binary variable. Yet the manner (e.g., mirroring conversation partners' language, using polite expressions, providing support and explanations, prompts and suggestions; Liu, Gerdtz, & Manias, 2016) and intensity (e.g., U. Fischer & Orasanu, 2000) in which people raise safety concerns varies. The ‘Walking the plank’ paradigm provides direct access to safety voice as speaking-up through the systematic observation of conversational processes in response to an observed risk, and novel insights may be drawn through conversation (Kendrick, 2017) or speech analysis (e.g., pitch; Kawahara & Morise, 2011; van Heuven & Boersma, 2001). We intend to test and present the linguistic nature of safety voice in future research.

Does personal relevance of harm shape safety voice?

We illustrated that safety voice can be observed in laboratory environments, and this enables the investigation of how the personal implication of physical harm shapes safety voice. The behaviour can emerge from individuals who are directly impacted by the consequences of a hazard (e.g., the person walking the plank) or observe others putting themselves in danger, and this may alter results. Different predictions exist for why people prevent harm to others versus oneself (Crockett et al., 2014), and these emerge from i) a stronger aversion to one’s own pain (i.e., economic exchange hypothesis), ii) an aversion to conflicting harm (i.e., guilt-aversion hypothesis), and iii) an equal evaluation of harm to self and others (i.e., empathy perspective). For example, Batson and colleagues found that perceiving someone as needing help increases empathic concern and helping (Batson et al., 2007), and our results suggested that victimhood does not influence safety voice. This opens up questions regarding the relationship between victimhood and safety voice, and whether it is explained by processes such as empathy and perspective taking. The ‘Walking the plank’ paradigm enables the investigation of these questions.

Can concepts for safety voice and harm-prevention behaviours be integrated?

Safety voice appears conceptually related to obedience to authority and the bystander effect, and the experimental paradigm may be used to investigate conceptual and operational overlap. Milgram’s behavioural study of obedience (Milgram, 1963) has been reconceptualised as operationalising an act of defiance (Miller et al., 1995) or resistance (Kaposi, 2017) that closely resembles safety voice (i.e., people repeatedly speak-up to resist a harmful order from an authority figure), and resistance may represent a special type of safety voice. Similarly, research on bystander interventions of people in need of assistance (Bennett et al., 2014; Darley & Latane, 1968; P. Fischer et al., 2011) operationalised harm-prevention behaviours and these have included non-verbal (e.g., walking to another room to intervene in sexual harrasment; P. Fischer et al., 2006) and verbal actions (van den Bos et al., 2009). However, the extent of overlap between these harm-prevention behaviours (e.g., raising a concern to a senior figure; Milgram, 1963) and triggers (e.g., the necessity of noticing and evaluating situations as dangerous; Latane & Darley, 1968) remains unaddressed. Evaluating overlap and integrating conceptualisations would provide an interesting research agenda, especially because, to our awareness, we are the first to indicate that unconcerned participants can step up to prevent harm, and the notion of 'unconcerned voice' would provide a novel take on long-established paradigms (e.g., disobedient participants might have objected about electric shocks on principle, not due to safety concerns).

Limitations

Three limitations of the experimental paradigm must be stated. That is, first, the paradigm cannot establish whether safety voice prevents physical harm: the paradigm presents a controlled hazard (i.e., the safety of its actual outcome is assured), and assumes that if the hazard were real physical harm would have been prevented. This is an important limitation that emerges from an ethical paradox: safety voice research aims to design safety interventions, but experiments cannot put participants in harm's way (e.g., actually breaking the plank, violence to participants, etc.). We illustrated how this can be addressed through manipulating risk perceptions that lead to safety concerns.

Second, the responses to the wrap-up questionnaire scales may be interpreted differently by those who raised a concern or remained silent, and may be understood as rationalising their behaviour (P. Fischer et al., 2006). For example, through obtaining safety concerns post-hoc, results may be interpreted as self-perception (Bem, 1967): people that remained silent were less likely to say they were concerned. This challenge is common for experiments that use a cover story to introduce a hazard, and we agree with Fischer and colleagues (2006) that the presentation of the questionnaire after the behaviour is the optimal procedure to collect additional data "without risking the credibility of [the] experimental design. Asking about these variables right after the danger manipulation and before measuring the dependent variable would have caused suspicion and unmasked our cover story" (p.272). This underscores the need for direct observations of safety voice, and the development of direct measures of its antecedents (e.g., physiological measures of safety concerns).

Third, the external validity of experiments is debated (Gigerenzer, 1984; Jiménez-Buedo & Miller, 2010) and because conclusions may not generalise (e.g., to the Intensive Care Unit, or mountaineering) the paradigm may thus not provide insights for unique environments. Yet, safety voice behaviours are highly contextual (i.e., they are shaped by antecedents and hazards), and, to enable conclusions on safety voice mechanisms, this calls for strict control over contextual variables through standardised assessments. That is, conclusions with high internal validity are near-impossible to draw in fast-paced environments with inconsistent presentation of antecedents and hazards, and mechanisms can only be established using highly standardised measures or scenarios presented across participants. Providing high internal validity, the proposed paradigm can isolate speaking-up in a standardised scenario and generalise with more certainty to contexts that have (manageable) characteristics tested through the paradigm, and external validity can be established through benchmark findings against other contexts.

Conclusion

Safety voice behaviours can be observed in laboratory experiments (and safety silence through assessing safety concerns). This is important because current safety voice methodologies have shortfalls, and experimental paradigms, despite their own limitations, are needed to address the behavioural nature of safety voice, reliance on memory and imagination, and relationship between safety voice and other variables. We presented the first experimental paradigm for investigating safety voice (the Walking the plank paradigm) that can address the requirements for safety voice experiments, and we illustrated how these can be evaluated. Through investigating safety voice experimentally, insight was provided on the importance of considering risk perception when interpreting behaviour, leading to a new two-dimensional typology for analysing safety voice behaviours. Our presentation of the paradigm adds to the debate on the need for appropriate methodologies for investigating harm prevention behaviours. The literature on safety voice has generated considerable insight into why people raise safety concerns, and the development of experimental methodologies advances the field: fostering the development of behavioural conceptualisations, new directions for research, and stronger interventions for the prevention of physical harm. People speaking-up about safety has saved countless lives, and experimentally examining the causes and nature of this behaviour has the potential to increase the prevalence and effectiveness for people to create safety through speaking-up.

CHAPTER 5:

A PRECISE BEHAVIOURAL CONCEPTUALISATION OF SAFETY VOICE AND SAFETY SILENCE

Preface

In the previous chapter I presented an experimental methodology for studying safety voice. This enables the investigation of gaps in the safety management and safety voice literatures, and it enabled a more nuanced and complete conceptualisation of safety voice as rooted in the extent to which people respond to hazardous situations (e.g., perceived risk, safety concerns). Consistent with systems perspectives on safety management (Leveson, 2002; Reason, 2000), I presented the first evidence that the act of safety voice is ecologically related to variables across levels of analysis (Chapter 3) and contingent upon the perception of risk (Chapter 4). This allowed the conceptualisation of safety voice as a two-dimensional behaviour based on the extent to which individuals engage in voice or silence and are unconcerned or concerned (see Figure 4.1 and Table 4.7). However, in order to understand the extent to which safety voice enables better safety management through the flow of safety-related information (Westrum, 2014), this concept requires further development and evaluation in terms of the extent to which people raise safety-related knowledge and motivations in speech (Christian et al., 2009).

In particular, a need exists to establish how the behavioural nature of safety voice may be scaled on a continuum (i.e., as a matter of degree to which people engage in speech) or as categorical variable (i.e., the extent to which safety voice manifests as distinct types of speech) in relationship to safety concerns, and how such a conceptualisation relates to characteristics of hazardous scenarios (e.g., how people relate to risk and others). For this, I opted to use the hazard in the Walking the Plank scenario alongside known antecedents (Barzallo Salazar et al., 2014; Burris, 2012; Duan et al., 2017; Tucker et al., 2008) because this provides a more optimal evaluation of the new behavioural model: it enables a comparison with known findings within an established paradigm and thus reduces uncertainty on how new findings should be interpreted (i.e., alternative findings may be attributed to the limited number of changes).

Thus, in Article 3, I evaluated the extent to which safety voice and silence can be scaled based on the distinct themes that people raise, and I evaluated this in relationship to safety voice antecedents and the temporal progression of hazards. By doing this, I propose and evaluate a conceptual model for the behavioural manifestation of safety voice and safety silence in terms of five types of speech (thematised around safety knowledge and motivation) that manifest differently in response to antecedents and over time (therefore contributing to research question 1: *What is the behavioural nature of safety voice?*). This enables systems perspectives on safety management (e.g., Leveson, 2002; Reason, 2000) to utilise more precise concepts for understanding how the lack of safety related information can contribute to accidents (Westrum, 2014).

Furthermore, I provide textual measures for assessing the extent to which people are concerned about safety and engage in different types of safety voice and safety silence (therefore contributing to research question 2: *What is the optimal way to investigate safety voice behaviour?*), and I indicate that, because safety voice is contingent upon risk perception, interventions are more effective for people that are more concerned about safety (therefore contributing to research question 3: *To what extent do interventions for promoting safety voice and reducing safety silence need to be tailored to the behavioural nature of safety voice?*).

In terms of terminology, I emphasise 'safety silence' because it provided for a consistent frame for the article (e.g., alignment of safety silence with reducing safety silence) and underscores the conceptual argument that safety voice and safety silence are phenomenologically strongly related in speech: safety silence manifests in less safety voice speech and not as its assumed absence (see Chapter 3) or binary opposite (see Chapter 4).

Study description

In this Chapter, I present two studies. First, I present a pilot study (n = 237) to establish the manipulations used within the article. This online study used a video-vignette of the Walking the Plank paradigm and supported the use of the manipulations. Second, in the article I present a laboratory study (n = 392) that used the Walking the Plank paradigm developed in Chapter 4. Through a textual analysis of participants' speech, I revealed five types of safety voice speech (based on safety knowledge and motivation) and indicate that these themes can be used to scale the extent to which people engage in safety silence. The proposed model advances the conceptual model for safety voice proposed in the previous chapters by conceptualising how (un)concerned voice and silence manifest in different degrees and types of speech (with the degree to which concerned individuals engage in safety silence being scalable in terms of five types of safety voice speech (i.e., informative, inquisitive, prohibitive, cautionary, and oblique safety voice; see Figure 5.1). Of significance for the prevention of harm, I revealed that conceptualising safety voice and silence in this manner is important because interventions only reduced specific types of safety silence (i.e., knowledge-based safety voice, not motivation-

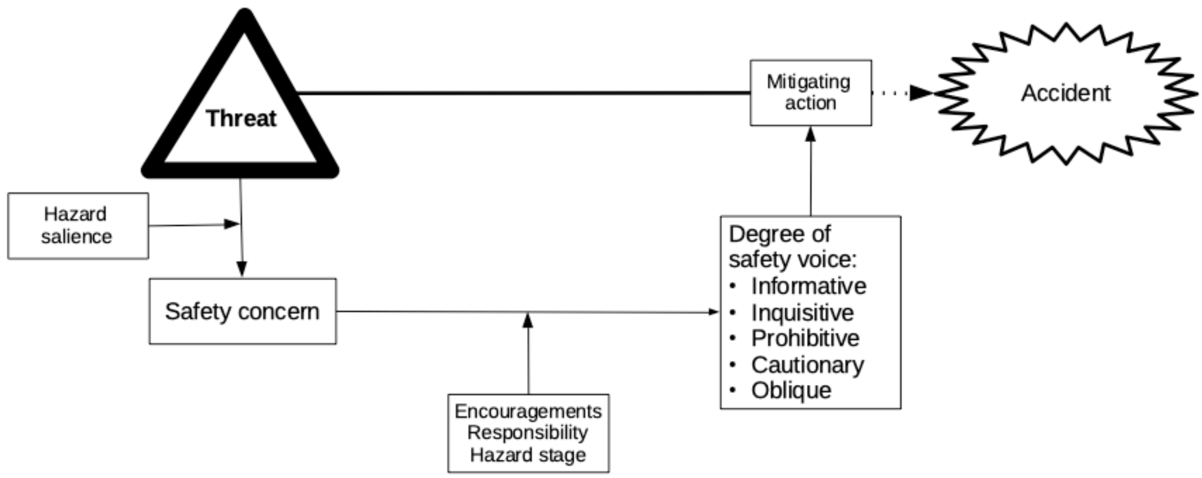


Figure 5.1. Threat Mitigation Model of safety concerns and degrees of safety voice.
Note: this model highlights that the dysfunctional momentum of threats towards accidents (Barton & Sutcliff, 2009) can be mitigated through safety voice that manifests as different degrees of knowledge- (i.e., informative, inquisitive) and motivation-based speech (i.e., prohibitive, cautionary), with antecedents impacting differently on concerns and speech.

based) and during particular stages of the hazard. Moreover, this revealed that the behavioural nature of safety voice is time-sensitive because the relationship between hazards and safety voice changes over time.

To enable the direct reproducibility of analyses and contribute towards open science, the study data and Jupyter notebook (version 3.7) have been submitted as supplementary material to Safety Science and a copy of the Jupyter notebook (converted to a printable format) is provided in Appendix E[25].

Authorship

I was responsible for designing and conceptualizing the study, data collection, formal data analysis and interpretation, and manuscript preparation (80%). Tom Reader and Alex Gillespie contributed to the conceptualisation of the study, refining the study design, interpreting results, writing Jupyter code and reviewing and editing the manuscript (20%). All data was collected at the Behavioural Research Lab at the LSE by research assistants executing the experimental protocol (Lindsie Arthur–Hulme, Nanne Houtsma and Alyssa Pandolfo). I was solely responsible for conceptualising, analysing and writing up the pilot study in this chapter. Further assistance was provided by Nelleke Noort-Van Dijk (video-editing the stimulus material) and Lindsie Arthur-Hulme (featuring in the stimulus material).

[25] Until the study data has been published as supplemental material to Article 3, I can be contacted to provide this.

Pilot study to Article 3

I conducted an online pilot study through Prolific (n = 237) to develop experimental manipulations for three variables known to shape safety voice: the awareness of hazards (Tucker et al., 2008), felt responsibility (e.g., Duan et al., 2017) and encouragements (e.g., Barzallo Salazar et al., 2014; Burris, 2012).

A 30-second video-vignette of the hazard in the Walking the plank paradigm was presented (i.e., two people; the partner walking the plank elevated to chair level)[26], preceded by information on the scenario presented in the video-vignette (i.e., you participate in a lab experiment; the plank used can only withstand 30kg/4.7 stone/66 pounds; only you are informed about the weight limit; your partner will test the plank as a footbridge). The manipulations were introduced independently alongside the vignette.

The hazard salience manipulation (n = 80; M_{age} = 30.000 years, SD_{age} = 9.626; $n_{students}$ = 31; n_{female} = 43; $n_{missing\_demographics}$ = 3) requested participants to evaluate a picture of a man talking on his phone whilst crossing a busy street and asked "What aspects of this picture make it a hazardous situation, where harmful outcomes might occur?" (salient condition), or: "What aspects of this picture make it a typical situation, one you could encounter any day?" (control condition).

The responsibility manipulation (n = 76; M_{age} = 30.620 years, SD_{age} = 10.475; $n_{students}$ = 33; N_{female} = 29; $n_{missing\_demographics}$ = 4) requested participants to consider: "Please think of a situation from your life where 'you' (clear condition)/ 'it was not clear who' (unclear condition) was responsible for the outcomes of the situation." Participants then briefly described the situation,

---

[26] A pilot study supported the use of the video-vignette: the proportion of concerned participants on the vignette (78%; 95CI: 65-92%) was not different from Chapter 3, study 3 (i.e., 68%), $t(36) = 1.513$, $p = .139$.

what they did, and how they felt. This manipulation involved situations external to the presented hazard to avoid potentially instructing participants to take charge and speak-up.

The encouragement manipulation ($n = 81$; $M_{age} = 30.923$ years, $SD_{age} = 10.536$; $n_{students} = 30$; $n_{female} = 37$; $n_{missing\_demographics} = 2$) described that the weight limit was provided by ‘the main researcher’ who, before walking out, introduced the participant to a research assistant that had not received any information about the plank. The research assistant stated one of two messages: ‘Please keep your thoughts and opinions to yourself. I do not like it when people share those, and I might then reduce your study reward because expressing your true feelings is not part of the task’ (discouraged condition). Conversely: ‘Please feel free to express your thoughts, and opinions. I like it when people share those, and it will not impact your study reward because expressing your true feelings is part of the task’ (encouraged condition).

Measures

Safety voice. A 5-point Likert scale (i.e., 1: Strongly agree; 5: Strongly Disagree) measured safety voice (‘I would have told her that walking the footbridge was a bad idea’). For binary calculation, the scale mid-point was coded as silence.

Safety concerns. A similar 5-point Likert scale measured safety concerns with three questions: ‘Walking the plank was dangerous’, ‘The plank was likely to break’ (for hazard salience), and ‘I was concerned about her walking the plank’ (for responsibility and encouragements). The inclusion of these items varied because the manipulations were introduced at separate time points, and this is highlighted in the results. The concern variables were recoded into binary variables, with the neutral scale-point (i.e., 3: unsure) coded as the active form of the variable (i.e., concerned, likely, dangerous).

Results and discussion

Findings (summarised in table 5.1) revealed that the three manipulations were effective in altering responses to a safety voice item (‘I would have told her that walking the footbridge

Table 5.1. Did the manipulations shape safety voice in the online pilot study?

		OR	Conclusion
Simple main effects			
Hazard salience		.566	no
Responsibility		.778	no
Encouragements		25.396*	yes
Modification of the effect of safety concerns			
Hazard salience	Control	4.928*	yes
	Salient	2.625	
Responsibility	Unclear	2.182**	yes
	Clear	1.132	
Encouragements	Encouraged	13.062**	yes
	Discouraged	2.187	

* $p < .05$; ** $p < .01$

was a bad idea'; the 5-point Likert scale with 1 = strongly agree was recoded with the scale mid-point as silence for binary calculation), but only as simple main-effect for the discouragement manipulation and as conditional effects for all manipulations on the predictive effect of safety concerns on safety voice (5-point Likert scale for: 'The plank was likely to break', 'I was concerned about her walking the plank'). This supported the use of the manipulations for Article 3.

To be more specific, *salient hazards* eliminated the effect of safety concerns on safety voice through a conditional effect. No simple effects were found for the influence of salient hazards on safety voice, $OR = .556$, $SE_{log\text{-}odds} = .490$, $p = .230$, or perceiving the plank as dangerous, $OR = .636$, $SE_{log\text{-}odds} = .554$, $p = .414$, and likely to break, $OR = 1.658$, $SE_{log\text{-}odds} = .507$, $p = .319$. Yet, and indicating hazard salience reduced uncertainty, participants in the control condition were 4.928 times more likely to speak-up if they were concerned (i.e., in terms of likelihood of breaking), $SE_{log\text{-}odds} = .777$, $p = .040$, not when hazards were salient, $OR = 2.625$, $SE_{log\text{-}odds} = .773$, $p = .212$.

Highlighting *clear responsibilities* moderated the effect of safety concerns on safety voice. Whilst the manipulation did not influence safety voice directly, $OR = .778$, $SE_{log\text{-}odds} = .566$, $p = .657$, or the extent that participants reported feeling responsible, $F(1,74) = 2.341$, $p = .130$, $\eta^2 = .031$, *observed power* = .327, stronger concerns only predicted safety voice when responsibilities were diffuse, *OR* = 2.182, $SE_{log\text{-}odds}$ = .428, *p* = .008, but no-longer when responsibilities were clear, *OR* = 1.132, $SE_{log\text{-}odds}$ = .470, *p* = .068.

Encouragements increased the likelihood of safety voice, especially when participants were concerned. Encouraged participants perceived the research assistant as more welcoming, $F(1,79) = 16.200$, $p < .001$, $\eta^2 = .170$, *observed power* = .978, and voiced more, *OR* = 25.28, $SE_{log\text{-}odds}$ = 1.40, *p* = .02. This interacted with the strength of safety concerns, *OR* = .35, $SE_{log\text{-}odds}$ = .50, *p* = .04, as holding safety concerns led to 55% more safety voice when encouraged, *OR* = 13.062, $SE_{log\text{-}odds}$ = .877, *p* = .003, but not when discouraged, *OR* = 2.187, $SE_{log\text{-}odds}$ = .778, *p* = .314.

Article 3:

The sounds of safety silence: interventions and temporal patterns reduce unique thematic absences of talk about safety

Submitted for publication in:

Safety Science (submission date: 25 June 2020).

Abstract

Research shows that withholding safety concerns on encountering hazards – *safety silence* – is a critical contributor to accidents. Studies therefore aim to prevent accidental harm through interventions for reducing safety silence. Yet, the behaviour remains poorly understood: it is unclear how it manifests in speech, may be scaled based on the degrees and types of safety voice (speaking-up about safety), progresses over time, and may be optimally reduced considering these. To address this, we proposed a conceptual model for the manifestation of safety silence and used a laboratory experiment (N = 404) to evaluate the implications for the effectiveness of three interventions (salient hazards, clear responsibilities, encouragements) across stages of a hazard. Results indicated that safety silence is scalable in terms of the degree that concerned people engage in five types of safety voice at different points in time, and we revealed this is important because interventions only reduce safety silence at unique hazard stages and for knowledge-based speech when people are concerned. This indicates that safety silence is situated and manifests as nuanced speech, and interventions are most effective when timed appropriately and people have safety concerns to speak-up about.

Keywords: safety voice; safety silence; harm prevention; intervention; speech; experiment.

Article highlights

- Safety silence is the act of withholding safety concerns about hazardous situation.
- Reducing safety silence is important because it can contribute to accidental harm.
- Yet, the behaviour remains poorly understood: interventions require new concepts and measures.
- Experimental data indicates safety silence manifests as five types of safety voice speech.
- This new typology is important: interventions and timing only reduce select types of speech.
- Evaluated concepts and measures enable new directions for research on safety voice and silence.

The words of the prophets are written on the subway walls, and tenement halls, and whispered in the sounds of silence.

– Simon & Garfunkel

Introduction

Safety silence is the act of withholding safety concerns about accidental harm (Schwappach & Richard, 2018; Tucker et al., 2008). In social and organisational settings (e.g., private transportation, intensive care units), the act of speaking-up about safety (termed 'safety voice') is recognised as crucial for mitigating hazardous conditions (Okuyama et al., 2014). However, people often do not engage in safety voice when encountering a hazard (Noort et al., 2019b). This is undesirable because safety silence has contributed to tragic outcomes in transportation (e.g., aerospace; Bienefeld & Grote, 2012; Cocklin, 2004; Moorhead et al., 1991; Tarnow, 1999), offshore oil drilling (Reader & O'Connor, 2014) and healthcare (Bromiley & Mitchell, 2009; Francis, 2013). Consequently, reducing safety silence is integral to improving organisational safety performance (M. A. Griffin & Neal, 2000; Hofmann et al., 2003) and an ethical and financial imperative (Novak, 2019).

Safety voice theory suggests that interventions can reduce the likelihood for safety silence (Noort et al., 2019b). For instance, through improving hazard salience (Tucker et al., 2008), people's felt responsibility (e.g., Duan et al., 2017) and leaders inclusiveness (e.g., Barzallo Salazar et al., 2014; Burris, 2012). Yet, it is unclear how interventions reduce safety silence because few insights exist regarding what concerned people do when they do not speak-up (e.g., speaking less overall, or only about safety), to what extent safety silence conveys clues that people are concerned, or the degree to which interventions target aspects of the behaviour (e.g., types, time-points). Presently, conceptual models recognise that safety voice is not a binary but may occur with different gradients (e.g., A. Jones & Kelly, 2014; Noort et al., 2019a) and distinct behaviours have been proposed for safety voice (e.g., explicit, respectful, oblique voice; Krenz et al., 2019; Pian-Smith et al., 2009) and silence (i.e., as withholding, absence of communication, or lack of involvement; Mumford, 2015). However, to our knowledge, this has not been theorised into a concepetual model for scaling types of safety voice and silence, and whilst measures have been developed (e.g., Krenz et al., 2019) it is unclear to what degree interventions reduce safety silence as a continuous (i.e., the degree and timing of speech) and categorical phenomenon (i.e., types of speech).

Thus, a need remains for models and measures that address how safety silence manifests, and for evaluating how interventions reduce safety silence. Without this, the field does not provide conceptual clarity or measurement of the behaviour it is trying to change, or recognise that operators may engage in a mixture of voice and silence behaviours (with their frequency, urgency, and timing shaping the strength and effectiveness of voice). Insighst on the behaviour's manifestation would enable the assessement of its quality and impact (Kolbe, Burtscher, et al., 2013). Therefore, through applying a validated experimental scenario for eliciting and observing the act of safety voice and silence (Noort et al., 2019a), we propose a conceptual model for how the behaviour of safety silence may be scaled and examine the degree to which safety silence is impacted by interventions and time.

Conceptualising safety silence

Safety silence refers to the act of withholding safety concerns during hazardous scenarios (e.g., Tucker & Turner, 2011) and is contrasted with *safety voice:* the act of raising safety concerns through discretionary verbal expressions (Conchie et al., 2012; Tucker et al., 2008). Although the literature focussing directly on safety silence is limited, it is implicit in research on communication and safety (Noort et al., 2019b). Due to its importance for safety management, safety silence is integral to behavioural models and measures of organisational safety (M. A. Griffin & Neal, 2000; Hofmann et al., 2003), safety culture and climate (Reader et al., 2015; Zohar, 2010), safety citizenship (Didla et al., 2009) and safety leadership (Barling et al., 2002). Furthermore, by virtue of safety silence involving the withholding of communication (e.g., reporting errors, advocating safe practice, transmitting warnings), insights from other voice concepts have been applied to conceptualise the effects of antecedents on safety silence (e.g., whistleblowing, upward dissent, employee voice and silence; Kassing, 2002; Morrison, 2014; Near & Miceli, 1985).

However, safety silence is different from other behavioural safety concepts (e.g., hand-washing, performing checklists) and voice/silence because it involves the withholding of concerns in any hazardous setting. Moreover, it extends to non-employees (e.g., patients reporting on deteriorating health, minibus passengers speaking-up about poor driving; Entwistle et al., 2010; Habyarimana & Jack, 2011). Therefore, safety silence requires a distinct conceptualisation to understand the types of behaviours that constitute the phenomenon, and their relationship to safety voice behaviours. Establishing this is important for reducing silence during critical incidents. For example, in high-risk setting safety silence is implicated in higher fatality rates, worse safety performance (Anicich et al., 2015; Kines et al., 2010), avoidable deaths (e.g., the Elaine Bromiley case; Fioratou, Flin, Glavin, & Patey, 2010) and aviation disasters (e.g., Tenerife collision, Air Florida 737, Swissair 111; Cocklin, 2004). This is because safety voice increases attention to safety (Kines et al., 2010), enables people to share observations and make concerns actionable (Barton & Sutcliffe, 2009), corrects and deters harmful actions (e.g., Palmer, 2016; S. H. Schwartz & Gottlieb, 1976; Sexton & Helmreich, 2003), and enables learning (Edmondson, 2003). Thus, safety silence is rooted in sensemaking about risk from hazards (Noort et al., 2019a; Schwappach & Gehring, 2014c).

Risk is important for conceptualising safety silence. Variation in risk perception (Slovic, 1987) can alter the interpretation of safety silence and may reduce the effectiveness of interventions. For instance, if people are unconcerned about hazards (e.g., falling ill from COVID-19), then silence indicates people had nothing to say and speech would be unrelated to safety. Conversely, for concerned people, silence would indicate that personal or contextual factors inhibited safety voice, with the extent that speech reflects safety themes indicating the strength of inhibiting factors. Safety silence is therefore not captured by the mere absence of voice (Brinsfield, 2013; van Dyne et al., 2003) or safety-related communication during hazardous situations (i.e., it is unclear whether concerns are withheld), and safety voice and safety silence may exist on a continuum. Few safety voice studies have investigated the content of speech and we argue that, in dynamic situations, people engage in speech about diverse safety themes, with their behaviour at different moments scaling from complete silence to explicitly raising the alarm.

The manifestation of safety silence

Psychological research on silence tends to conceptualise the phenomenon as the absence of verbalisation (Valle, 2019), talking less about a topic (e.g., risk; Mumford, 2015), or veiling the withholding of a topic through other speech (Morison & Macleod, 2014). Different linguistic types of silence have been conceptualised such as acoustic (i.e., no speech) and thematic silence (Kurzon, 2007, 2011). However, theoretical insights on the nature of *safety* silence are ambiguous because the literature has not clarified its manifestation and conceptualised it inconsistently as “a bounded act of withholding, a more general state of absence of communication, and a lack of being involved or active" (Mumford, 2015; p.34). For instance, distinct manifestations of safety silence are not addressed through statistics representing the extent of safety voice, post-hoc statements on silence, or available measures for safety voice and silence (Manapragada & Bruk-Lee, 2016; Tucker & Turner, 2011). An exception to this are the different types of silence proposed (van Dyne et al., 2003) and established (Brinsfield, 2013) for employee silence such as defensive, deviant and relational silence. However, these pertain to motives for withholding voice, and do not scale how silence manifests in terms of the content people raise. We distinguish motives from actions because identical motives (e.g., the desire to prevent harm) may lead to different utterances (e.g., 'please be careful', 'stop doing that') and are not necessarily expressed in content (e.g., when sharing safety limits). Research on aircrew conversations provides another notable exception (e.g., co-pilots providing hints or questioning captains; U. Fischer & Orasanu, 2000; Sassen, 2005), but no model exists for the manifestation of safety silence in speech. Thus, research suggests safety silence may manifest in distinct types, but insights, data and measures remain limited. Additionally, it is uncertain whether insights on the manifestation of safety voice can be applied to the manifestation of safety silence. Studies have rarely conceptualised and operationalised safety silence in terms of the withholding of safety concerns (Noort et al., 2019a) and may have confounded concerned and unconcerned participants. There have been few attempts to investigate the degree to which antecedents reduce safety silence, or how this varies over time. This appears critical: without this, the effectiveness of interventions for reducing safety silence (e.g., providing encouragements; Barzallo Salazar et al., 2014) remains uncertain.

Thus, we argue that safety voice and safety silence are not dichotomous: in dynamic safety critical scenarios concerned individuals may discuss a range of safety-related themes, with the manifestation of these in speech determining the quality and strength of safety voice and silence over time (Kolbe, Burtscher, et al., 2013). Conceptualising this is needed for disentangling distinct manifestations of the behaviour, and for evaluating interventions. Yet, to date, there is no conceptual model and little empirical data on how safety silence manifests. Through using an experimental paradigm and analysing speech, in the current study we aim to contribute a conceptual model for how safety silence can be scaled in relation to safety voice, and concrete indicators that can be used to test the effectiveness of interventions for reducing safety silence at different timepoints.

Current study: A behavioural investigation to scale safety silence

Here, we investigate how safety silence may be scaled in terms of safety voice speech, and the implications of this for interventions, in an experimental scenario. The 'Walking the Plank' paradigm presents participants with an apparent safety problem (a research assistant walking across a 'weak' and elevated wooden plank; Noort et al., 2019a) that can only be mitigated through raising safety concerns. Because the scenario measures hazard perceptions, it enables the assessment of safety silence behaviours (Noort et al., 2019a). Through using this paradigm, to our knowledge the only one of its type, we can directly observe behaviour, test a model for scaling safety silence and evaluate when interventions are successful at reducing safety silence.

Scaling safety silence

Our first research question investigates the degree to which safety silence can be scaled based on safety voice speech. Theory suggest that, during safety-critical scenarios, people can say nothing (i.e., acoustic silence; Kurzon, 2011), engage in unrelated speech (i.e., veiled and thematic silence; Kurzon, 2011; Morison & Macleod, 2014), or raise concerns. We suggest that, when concerns are fully withheld, it appears self-evident that this may appear as no (acoustic) or unrelated speech (thematic silence). Conversely, safety voice may be understood as the strongest way to express concerns in speech for concerned participants. Yet, importantly, people may only partially withhold safety concerns and produce some, but less, meaningful communication on safety that is not captured by binary concepts (A. Jones & Kelly, 2014) because the degree to which specific themes (e.g., safety information, the desire to avoid harm) feature in conversations can vary. Safety silence may therefore manifest in speech as a continuous (i.e., the degree of safety voice speech) *and* categorical phenomenon (i.e., types of safety voice speech). Thus, during hazardous scenarios people characterised as engaging in safety silence may engage in some safety voice, and we propose that safety silence should be conceptualised as the degree to which the withholding of safety concerns is manifested in speech that expresses distinct safety themes.

Specifically, we propose that safety silence can be scaled based on distinct types of safety voice related to safety knowledge and safety motivation. Initial evidence exists for distinct ways to raise concerns (e.g., respectful, explicit, oblique; Friedman et al., 2015; Kassing, 2002; Krenz et al., 2019; Park et al., 2013; Pian-Smith et al., 2009) and whilst a unified model remains absent these may be used for conceptualising and scaling safety silence. In particular, we propose that because i) beliefs and intentions provide the content of communication (Searle, 2008), and ii) because safety knowledge and motivation shape safety participation behaviours such as voice (Christian et al., 2009), this should manifest in distinct speech. For instance, encountered hazards can prompt different perceptions about safety (e.g., uncertainty on safety limits, concern for others' wellbeing; the content of safety voice), and if people discuss safety concerns with others they make sense of perceived risks and evaluate intentions to avoid harm through safer action (Brinsfield, 2013; Gruman & Saks, 2014; Searle, 2008; Turner & Gray, 2009). Thus, we propose that when people withhold safety concerns, speech reflects less

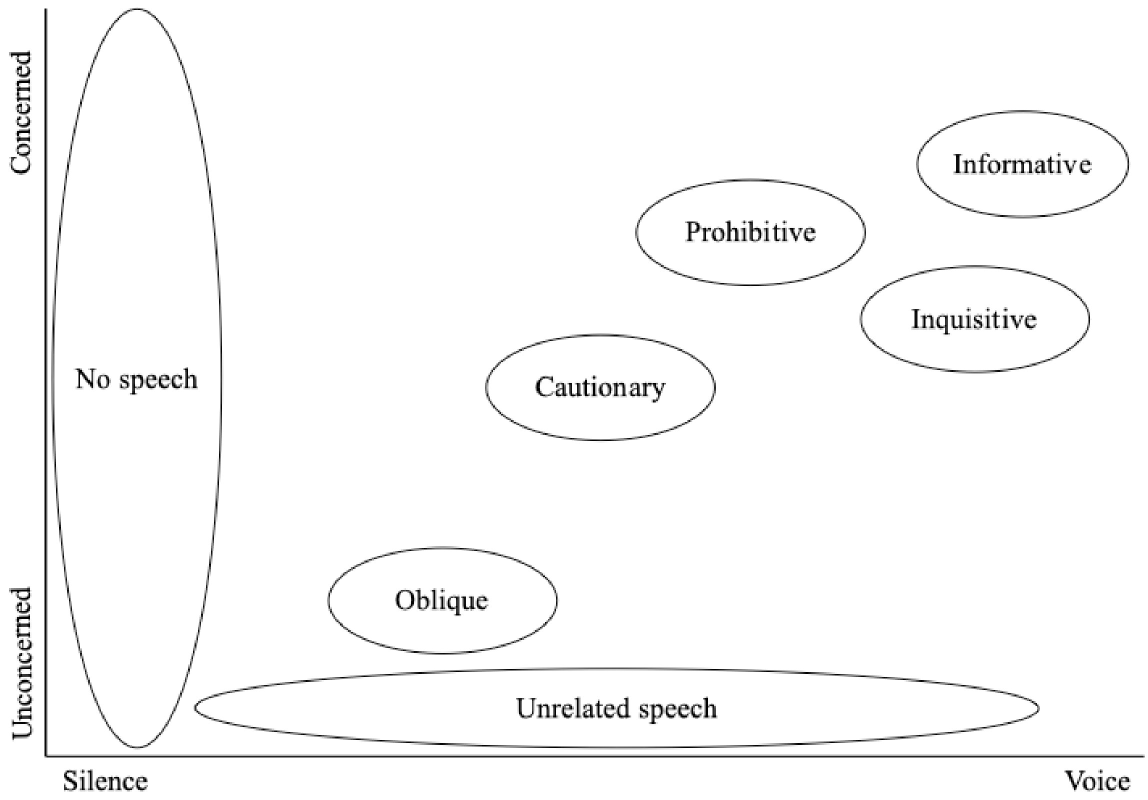


Figure 5.2. Model for the manifestation of (un)concerned voice and silence in speech. *Note: coordinates for the five types of safety voice reflect correlations with safety voice and safety concern dictionaries, with the areas for no and unrelated speech reflecting these can occur for all degrees of concerns and voice, respectively.*

discussion about safety knowledge and motivations, and we suggest this appears as five types of safety voice speech (see Figure 5.2).

First, through raising safety concerns people can discuss safety information. When people warn others they declare their safety beliefs (Searle, 2008) and evidence indicates that people raise safety concerns through appealing to facts or better solutions (Kassing, 2002), logical arguments (Schwappach & Gehring, 2014b), or, conversely, requesting clarifications (Pian-Smith et al., 2009). This enables people to make sense of the nature of anticipated or encountered hazards (Weick, 2010) and evaluate appropriate actions. Thus, we expect that safety silence may manifest as less discussion of safety knowledge through less provision of safety information (i.e., informative safety voice) and fewer requests for clarifications (i.e., inquisitive voice).

Hypothesis 1a: Safety silence manifests as less informative safety voice.

Hypothesis 1b: Safety silence manifests as less inquisitive safety voice.

Second, through raising concerns people can express the desire for intended states of the environment (eg., taking action, avoiding harm; Searle, 2008). Because safety motivations lead to safety voice (Christian et al., 2009), we propose that safety silence should be reflected in less speech that clarifies the intention to avoid harm such as prohibitive statements (e.g., 'please, stop that'), or tentative cautionary statements (e.g., 'be careful'). Prohibitive (i.e., explicit, blunt) statements enable clarity on the desire to avoid harm (Krenz et al., 2019) through crisply advocating for different actions (Pian-Smith et al., 2009) or threatening with resignation (Kassing, 2002). Cautionary statements express a desire to avoid harm whilst conveying more respect (Krenz et al., 2019). Because the motivation for avoiding harm reduces safety silence (Christian et al., 2009), this may be reflected in speech and safety silence may therefore involve less prohibitive and cautionary safety voice.

Hypothesis 1c: Safety silence involves less prohibitive safety voice.

Hypothesis 1d: Safety silence involves less cautionary safety voice.

Finally, people are indicated to raise concerns through unclear utterances (e.g., 'okay', 'ha?', 'hmm?', joar?', 'how?'; Krenz et al., 2019), and oblique speech (Pian-Smith et al., 2009) merely hinting at concerns held (U. Fischer & Orasanu, 2000). These utterances provide unclear content and thus the relationship with expressing safety knowledge and safety motivation is not straightforward. This manifestation of safety silence may emerge because the hesitancy to express safety concerns (e.g., due to higher cost of speaking-up) leads to mitigated speech (Edmondson, 1999; U. Fischer & Orasanu, 2000) that manifest in partial statements or utterances that are not explicit but imply concerns in-situ. Thus, we finally expect that safety silence may manifest in oblique speech.

Hypothesis 1e: Safety silence involves less oblique safety voice.

Reducing safety silence

Our second research question investigates the degree to which interventions reduce safety silence. Safety voice behaviour is attenuated (e.g., in occurrence, assertiveness of communication, repetition, explicitness) by situational variables (e.g., leadership styles, national culture; Barzallo Salazar et al., 2014; Rhee et al., 2014; Weiss et al., 2018) and it varies in effectiveness for how individuals and groups (e.g., safety managers, flight crews, operating teams) understand and decide on safety (e.g., problem-solving, being listened to; A. Jones & Kelly, 2014; Orasanu & U. Fischer, 1992). This means that it is important to investigate the relationship between situational variables and the manifestation of safety silence in order to design effective interventions (Noort et al., 2019a). Yet, few studies have directly observed safety voice whilst manipulating interventions (Barzallo Salazar et al., 2014; Friedman et al., 2015; Hodges, 2018), and we are not aware of studies that manipulated interventions whilst assessing variation in participants' safety concerns and speech. Without assessing this, studies i) assume that hazards elicit concerns, ii) confound concerned and unconcerned participants and iii) may not reduce the active withholding of safety concerns but increase the perception of risk. Accordingly, a need exists to evaluate the degree to which interventions can reduce safety silence.

We propose that interventions for reducing safety silence work optimally in terms of manifesting in more safety voice speech when participants are more concerned. Research using the Walking the Plank paradigm has indicated that safety silence is associated with participants reporting they are unaware of hazards, feel less responsible and worry less about the consequences of speaking-up (Noort et al., 2019a). This is consistent with proposed interventions for hazard salience (Tucker et al., 2008), felt responsibility (e.g., Duan et al., 2017) and encouragements (e.g., Barzallo Salazar et al., 2014; Burris, 2012). Through applying these manipulations, we enable the evaluation of our conceptual model against the literature.

First, safety voice is associated with people being aware (Lindberg et al., 2013; Manias, 2015) and concerned about hazards (Gurung et al., 2017; Manapragada & Bruk-Lee, 2016; Schwappach & Gehring, 2014a, 2014c). We suggest this leads to reduced safety silence because salient hazards (e.g., reminders of death) elicit risk perceptions through increasing perceived threat and uncertainty (i.e., outcomes are not clear a-priori; Burke et al., 2010). Theory on risk communication and uncertainty management suggests that uncertainty can be managed through information-sharing (e.g., speech) that creates shared awareness, (dis)confirms risk perceptions and evaluates appropriate actions (Brashers, 2001; Lindell & Perry, 2012). Increasing hazard salience should therefore manifest in less safety silence.

Hypothesis 2a: Salient hazards reduce safety silence when people are concerned.

Second, ample research has indicated that felt responsibility for situational outcomes increases voice (Aydon et al., 2016; Bickhoff et al., 2016; Duan et al., 2017; Jackson et al., 2010; Lyndon, 2008; Malvey et al., 2013; Manias, 2015; Nembhard, Yuan, et al., 2015; Schwappach & Gehring, 2014a). This is because clear responsibilities increase the intention to communicate in order to i) decide on appropriate action (P. Fischer et al., 2011; Lindell & Perry, 2012; Weiss et al., 2018), ii) redefine optimal performance (Fuller et al., 2006), and iii) explicitly prevent harmful outcomes (Weiss et al., 2014). Clear responsibilities describe the accountability for situational outcomes (e.g., harm) and increase the willingness to accept accountability for future consequences (Fuller et al., 2006). This may legitimise the sharing of safety knowledge through group norms for communicating risk. Thus, safety silence may be reduced through increasing the extent people feel responsible for the outcomes of hazardous situations.

Hypothesis 2b: Felt responsibility reduces safety silence when people are concerned.

Third, encouragements can communicate favourable norms for speaking-up. Research indicates that people speak-up more to receptive leaders (e.g., through transformational leadership styles; Bickhoff et al., 2016; Nembhard & Edmondson, 2006). This is because explicit communication is more likely when others are supportive (Brashers, 2001; Lindell & Perry, 2012) and the costs of safety voice are low (Edmondson, 1999; P. Fischer et al., 2006; Lindell & Perry, 2012). Supporting this, encouraged participants are shown to be more likely to speak-up (Barzallo Salazar et al., 2014). Thus, we expect that safety silence is reduced through providing encouragements.

Hypothesis 2c: Encouragements reduce safety silence when people are concerned.

We proposed that safety silence can manifests in speech as a continuous (i.e., degrees of speech) *and* categorical phenomenon (i.e., types of speech). This suggest that interventions may only reduce safety silence for specific manifestations. Because insights remain scant, we explore how interventions impact on reducing specific types of safety silence. Arguably, safety silence may be reduced most in terms of speech related to safety knowledge (i.e., inquisitive and informative safety voice). This is because hazard salience, felt responsibility and encouragements involve clarity on safety information and norms for communicating this. Through exploring this we aim to reveal whether interventions for reducing safety silence should be tailored to types of safety silence.

The effect of time on reducing safety silence

Our third research question investigates the degree to which safety silence manifests differently over time. Time provides a natural influence on safety silence, yet few studies have conceptualised temporal differences in safety silence, or the effect of interventions across stages of hazardous scenarios. An exception, Farh and Chen (2018) showed that intervention success depends on intervention timing (i.e., preparation versus execution of procedures). This indicates safety silence may manifest differently across stages of hazardous scenarios, with interventions targeting distinct aspects of safety silence. We suggest that, in temporal order, hazardous scenarios may i) be anticipated as a potential future state (e.g., designing new systems, planning routes), ii) be physically encountered (e.g., medical alarms sounding), and iii) provide the potential for imminent harm (i.e., initiated actions with impending outcomes). In the first two stages, harm is not immediate and remains distal compared to initiated actions that require immediate action. Arguably, this may elicit more conceptual evaluations (i.e., knowledge-based speech) for the early phases of hazards, and more discussion of the intention to avoid harm for later stages (i.e., motivation-based speech).

Hypothesis 3: As hazardous scenario progress, safety silence is manifested in less knowledge-based speech and more motivation-based speech.

In addition, we explore intervention effects over time. Little evidence exists that enables explicit hypotheses, but because hazard salience, felt responsibility and encouragements involve clarity on safety information and norms for communicating, the interventions may be more effective for reducing safety silence during the early stages of the hazard.

Method

Design

Within a laboratory environment, participants engaged in the validated Walking the Plank paradigm (Noort et al., 2019a). Under the guise of a creativity study, this paradigm presented an apparent hazard of walking a footbridge (i.e., the plank supposedly only held 30kg), and enabled the direct observation of safety silence in response to controlled hazards.

The protocol had three stages. First, after obtaining informed consent, participants engaged in a 5-minute creativity task where they described the possible uses of a plank and four blocks of wood. Second, they engaged in task with a research assistant to test the feasibility and creativity of the ideas of a ‘previous participant’ (i.e., a standard set: shelving, mirror, juggling, footbridge, piece of art). Finally, they completed a questionnaire and were fully debriefed. For the footbridge idea, the protocol required the research assistant to i) introduce the footbridge idea (“Hmm. This idea is pretty obvious, but I haven't seen it before. Could you build a footbridge, please?"), ii) prompt the participant to place the plank across two chairs, ii) state the intention to walk the plank ("I will now test the footbridge idea by walking over it"), and iii) walk the plank (stepping onto the footbridge at one chair, stepping off the footbridge at the other).

An online pilot study using a video-vignette of the Walking the Plank paradigm (n = 237) indicated that hazard salience, $OR = 4.928$, $SE = .777$, $p = .040$, responsibilities, $OR = 2.182$, $SE = .428$, $p = .008$, and encouragements, $OR = 13.062$, $SE = .877$, $p = .003$, modified the effect (Vanderweele, 2009) of safety concerns on safety voice. Additionally, encouragement had a direct effect on safety voice, $OR = 25.396$, $SE = 1.398$, $p = .021$. These manipulations were therefore introduced within the protocol.

For the hazard salience manipulation, participants evaluated a picture of a man talking on his phone whilst crossing a busy street, and were asked "What aspects of this picture make it a hazardous situation, where harmful outcomes might occur?" (salient condition), or: "What aspects of this picture make it a typical situation, one you could encounter any day?" (control condition). For the responsibility manipulation, participants read: "Please think of a situation from your life where 'you' (clear condition)/ 'it was not clear who' (unclear condition) was responsible for the outcomes of the situation." Participants then described the situation, what they did, and how they felt. For the encouragement manipulation, the research assistant stated one of two messages: 'Please keep your thoughts and opinions to yourself. I do not like it when people share those, and I might then reduce your study reward because expressing your true feelings is not part of the task' (discouraged condition). Conversely: 'Please feel free to express your thoughts, and opinions. I like it when people share those, and it will not impact your study reward because expressing your true feelings is part of the task' (encouraged condition). The hazard salience and responsibility manipulations were presented electronically in counter-balanced order before the creativity task. The encouragement manipulation was introduced by the research assistant before the 'previous participants ideas' were tested. The eight conditions were randomised across all participants and research assistants were blind to the study hypotheses.

Participants

404 participants ($n_{students} = 377$; $n_{female} = 277$, Age $M_{(sd)} = 22.897_{(5.386)}$, $n_{missing\_demographics} = 9$) consented to participate (including anonymised data to be archived and used within public domains), completed the study and were rewarded for their time. The full dataset is available as supplementary material. Twelve participants were excluded from analyses (i.e., 10 technical issues with video recording, 2 non-responses to whether the scenario elicited concerns). Participants spoke fluent English with 97% being native speakers (n = 166) or speaking English for more than five years (n = 216).

Measures

Measures included self-report and behavioural measures tailored to the laboratory environment (for an overview, see Table 5.2).

Safety concerns. Safety concerns were measured with a 5-point Likert scale item: 'I was concerned about the footbridge idea'. To enable the identification of the continuous safety concern dictionary the item was adapted to concerned (i.e., ≥ 3) and unconcerned (i.e., ≤ 2). The concern dictionary scored the frequency of concerned words for participants' speech.

Safety silence. Safety silence was coded based on transcribed video-recordings of the hazardous scenario (i.e., introduction of the footbridge idea up to moving on to the next phase of the study). Safety voice was coded when participants verbally indicated they were concerned about the research assistant walking the plank (i.e., a risk was indicated, the situation prohibited, proceedings questioned, caution urged, or a concern suggested through an oblique expression). Otherwise the behaviour was coded as 'not safety voice', $ICC(1,1) = .749$, $p < .001$, and labelled 'safety silence' when participants held a safety concern. Seven participants that withdrew their voice (i.e., they spoke-up, but backtracked and allowed the footbridge to be walked) were coded as safety voice because an option to respond was given. Conversely, independent conversational gasps and apologies were not considered safety voice.

To measure safety silence and safety voice in speech, participant text was scored with i) LIWC2015 dictionaries for risk, perceptions, future-orientation, personal pronouns, negation and formalities (Pennebaker et al., 2015), ii) the communication vagueness scale (Hiller et al.,

Table 5.2. Overview of manipulations and measurement characteristics for study variables.

Characteristic	Detail	Levels/Scale
Protocol		
Manipulations	Hazard salience	Salient/control
	Responsibility	Clear/unclear
	Encouragements	Encouraged/discouraged
Self-report measures		
Safety voice	'I told the RA that walking the footbridge was a bad idea (before the RA walked on it)'	5-point Likert scale (1: Strongly disagree)
Concerns	'I was concerned about the footbridge idea'	5-point Likert scale (1: Strongly disagree)
Felt responsibility	'I would feel obligated to raise any concerns I had'	5-point Likert scale (1: Strongly disagree)
Perceived social risk	'I felt the RA was very unfamiliar to me'	5-point Likert scale (1: Strongly disagree)
	'I felt the RA might lead me to compare badly to him/her'	
	'I felt the RA might reduce my influence over the situation'	
	'I felt the RA might cause me to have strong negative feelings'	
	'I felt the RA might exploit me'	
	'I felt the RA might bring out the worst in me'	
	'I felt the RA had a good understanding of the situation'	
Behavioural measures		
Safety silence	Observed behaviour in video-observations	(Not) safety voice (0/1)
	Safety voice dictionary	Number of uttered
	Informative	
	Inquisitive	
	Prohibitive	
	Cautionary	
	Oblique	
Concerns	Safety concern dictionary	Number of uttered
	Disfluencies dictionary	
Immediacy of harm	Observation of safety voice occurring during the footbridge idea	Four time-points (1: introducing the footbridge)

1969), and iii) safety voice dictionaries (i.e., informative, inquisitive, prohibitive, cautionary, oblique; see Table 5.3). Safety voice dictionaries were developed through identifying words associated with coded safety voice behaviour, identifying synonyms using word vectors (Mikolov et al., 2013) and manually evaluating patterns through author discussion. Dictionary scores (i.e., continuous scales) are therefore distinct from observations (i.e., binary scales).

Hazard stages. The timing of safety silence was assessed using four time-points (i.e., introducing the footbridge idea; placing the plank across the chairs; stating the intention to walk the plank; the research assistant stepping off the plank). Concerns raised after the research assistant walked the plank were not considered safety voice (i.e., harm could no longer be prevented). The timing was coded with 'substantial' interrater reliability (*AC1* = .737, 95CI: .622 – .853).

Questionnaire items. Felt obligation was measured with an adapted survey item ('I would feel obligated to raise any concerns I had'; Liang et al., 2012) and six items measured social risk (α = .762; Noort et al., 2019a).

Table 5.3. Description of safety concern and safety voice dictionaries.

Dictionary	Definition	Word list
Safety concern dictionaries		
Concerns	Speech indicates a concern is held	"actually", "although", "because", "believe", "but", "could", "doubt", "expect", "feel", "for", "guess", "hopefully", "if", "just", "know", "maybe", "mean", "not", possibly", "perhaps", "pretty", "probably", "rather", "really", "still", "suppose", "sure", "think", "though", "thought", "too", "uhm", "quite", "yet", "seems", "unless", "wonder"
Disfluencies	Interruptions in fluent speech	"ah", "ahh", "ahhh", "ahhhh", "duh", "dunno", "eh", "ehh", "er", "erm", "errr", "gee", "geez", "hah", "haha", "hahah", "hahaha", "hahahah", "hahahaha", "heh", "hehe", "heheh", "hehehe", "hey", "hmm", "hmmm", "hmmmm", "hmmmmm", "huh", "humm", "nah", "nevermind", "nope", "oh", "ohh", "ohhh", "ooh", "oooh", "uh", "uhh", "uhhh", "uhm", "uhum", "umm", "ummm", "ummmm", "whoa", "yah"
Safety voice dictionaries		
Informative	Informing the other about hazards, outcomes or safe alternatives	"4.7", "almost", "animal", "because", "bend", "break", "but", "child", "could", "dangerous", "enough", "even", "fall", "for", "harm", "hold", "hurt", "just", "kg", "kid", "kilo", "kilograms", "kilos", "least", "less", "limit", "load", "maximum", "much", "not", "only", "pounds", "restriction", "said", "says", "six", "sixty", "small", "so", "someone", "stable", "still", "stone", "support", "sustain", "take", "than", "thirty", "up", "very", "weigh", "weight", "when", "work", "would"
Inquisitive	Requesting hazard-related information from the other	"?", "before", "did", "do", "gon", "have", "na", "to", "want", "wanted", "would", "yeah", "you", "are", "ask", "how", "sure", "understand", "why", "when", "who", "what", "where", "will", "really"
Prohibitive	Ending the unfolding hazard through explicitly indicating risk or a need to stop action	"danger", "safe", "feasible", "quite", "fine", "good", "depends", "work", "exactly", "too", "n ot", "won't", "'m not", "not gon", "can't", "n't", "try", "step", "test", "I", "me", "no no", "wa it", "should", "wanted", "stop", "able", "afraid", "bad", "better", "keep", "me", "mind", "m ust", "need", "no", "unless", "too", "wrong", "work"
Cautionary	Urging others to take care in dealing with the hazard	"careful", "move", "let me", "wait", "too", "need", "help", "hand", "simply", "just", "keep", "should", "take", "try", "better"
Oblique	Hinting at holding a negative evaluation of the hazard	"!", "ah", "alright", "anyway", "damn", "didn", "god", "gon", "gosh", "great", "guess", "ha", "haha", "hahaha", "heck", "hell", "hey", "hmm", "hmmm", "huh", "joar", "kidding", "lol", "oer", "ok", "okay", "oof", "ooh", "oohoo", "oops", "ow", "phoe", "phoo", "say", "shit", "sorry", "uh", "whoo", "whoops", "wow", "brave"

Results

Analyses were conducted with Python 3.7 (using the pandas, numpy, scipy, statsmodels, spacy and scattertext packages). The Jupyter notebook and supporting files are provided as supplementary material. Accordingly, to improve readability, statistics are summarised, and non-significant statistics are presented as *'ns'*.

Manipulation and dictionary checks

Manipulation checks indicated the scenario and experimental manipulations worked as intended, with mixed success for the responsibility manipulation.

Scenario. The scenario elicited safety concerns for 78.8% (95CI: 74.8 – 82.9%) of the participants, $t(391) = 38.153$, $p < .001$, and 47.7% (95CI: 42.7 – 52.7%) raised a safety concern, $t(391) = 18.886$, $p < .001$. Excluding unconcerned participants did not alter this: 50.2% (95CI: 44.6 – 55.8%) of concerned participants spoke-up, $t(308) = 17.607$, $p < .001$. The distribution of (un)concerned safety voice and silence is presented in Table 5.4.

Participants uttered 18,078 words (M = 46.117; SD = 37.559), with participants raising concerns uttering more words (M = 64.433; SD = 39.576) than those not raising concerns (M = 29.410; SD = 26.226), $F(1,390) = 108.388$, $p < .001$, $\eta^2 = .217$.

Manipulations. Hazard salience led to stronger concerns than the control condition, $F(1,390) = 4.871$, $p = .028$, $\eta^2 = .012$: a 1.78 times (95CI: 1.081-2.924) higher likelihood that participants were concerned, $SE = .254$, $p = .023$. The responsibility manipulation

Table 5.4. Distribution of voice and silence.

	Concerned		**Unconcerned**		**Total**	
	n	$\%_{(SE)}$	n	$\%_{(SE)}$	n	$\%_{(SE)}$
Voice	155	$50_{(2.8)}$	32	$39_{(5.4)}$	187	$48_{(2.5)}$
Silence	154	$50_{(2.8)}$	51	$61_{(5.4)}$	205	$52_{(2.5)}$
Total	309	$79_{(2.1)}$	83	$21_{(2.1)}$	392	$100_{(-)}$

Percentages total 100% within a column, except for the total of (un)concerned. (Adapted from Noort et al., 2019b) unexpectedly did not increase felt responsibility, $F(1,390) = .318$, $p = .573$, $\eta^2 = .001$. Yet, suggesting participants in the responsibility condition felt a shared obligation with the research assistant, participants uttered more 'we', $F(1,390) = 5.066$, $p = .025$, $\eta^2 = .001$. Encouraged participants perceived less social risk from the research assistant, $F(1,387) = 16.677$, $p < .001$, $\eta^2 = .041$, and uttered more words, $F(1,390) = 6.504$, $p = .011$, $\eta^2 = .016$. Furthermore, their speech was more informal, self-focussed and less negated, $F(1,390)$s ≥ 4.181, ps $\leq .042$, η^2s $\geq .011$.

Dictionary validation. Providing a novel measure for safety concerns, whether participants were concerned accurately related to dictionary scores for safety concern, $F(1,390) = 4.446$, $p = .036$, $\eta^2 = .011$. Indicating a possible tension between raising safety concerns and perceiving social risk, concerned participants' speech was more disfluent, $F(1,390) = 5.574$, $p = .004$, $\eta^2 = .022$. Safety voice behaviours were coded with 'substantial' or better interrater reliability (Wongpakaran et al., 2013) in terms of whether it involved informative ($AC1 = .768$, 95CI: .633 – .903), inquisitive ($AC1 = .959$, 95CI: .911 – 1.007), prohibitive ($AC1 = .837$, 95CI: .731 – .944), cautionary ($AC1 = .862$, 95CI: .767 – .957), and oblique speech ($AC1 = .688$, 95CI: .536 – .839). After identifying synonyms, the dictionaries related accurately to the intended behaviour (e.g., informative versus not informative), $F(1,390)$s ≥ 26.169, ps $< .001$, η^2s $\geq .063$. Dictionaries provided one composite safety voice dictionary, and this accurately distinguished between participants observed to voice or remain silent, $F(1,390) = 138.085$, $p < .001$, $\eta^2 = .261$. The safety voice dictionary was associated with self-reported safety voice, $r = .490$, $p < .001$, and only with the concern dictionary, $r = .813$, $p < .001$, not self-reported concerns, $r = .077$, $p = .126$. Means and correlations for study variables and manipulations are presented in Table 5.5.

Table 5.5. Spearman correlations, means and standard deviations of variables.

		M	SD	1	2	3	4	5	6	7	8	9	10	11	12
1	Salient hazard	.49	.50												
2	Clear responsibility	.50	.50	.015											
3	Encouraged	.49	.50	.040	-.015										
4	Safety voice dictionary	16.86	15.33	.014	-.036	.144**									
5	Concerned dictionary	3.80	4.60	.035	-.040	.136**	.813***								
6	Disfluencies dictionary	1.83	1.79	-.015	-.014	.054	.384***	.337***							
7	Informative dictionary	6.78	7.93	.027	-.032	.149**	.905***	.803***	.263***						
8	Inquisitive dictionary	5.86	5.41	.046	-.015	.121*	.897***	.687***	.323***	.729***					
9	Prohibitive dictionary	3.30	3.95	-.009	-.091	.131**	.849***	.744***	.326***	.723***	.715***				
10	Cautionary dictionary	.69	1.06	-.031	-.030	.069	.557***	.468***	.310***	.458***	.487***	.518***			
11	Oblique dictionary	1.77	1.89	-.015	-.051	.075	.379***	.174***	.434***	.206***	.281***	.232***	.249***		
12	Wordcount researcher	69.63	22.24	-.043	-.066	.017	.306***	.239***	.429***	.169***	.317***	.298***	.216***	.281***	
13	Wordcount participant	46.12	37.56	-.011	-.021	.135**	.951***	.819***	.456***	.863***	.861***	.802***	.553***	.357***	.347***

N = 392; * *p*<.05; ** *p*<.01; *** *p*<.001

Scaling safety silence

Supporting hypotheses 1a-e, participants engaging in safety silence uttered words (M = 29.838, SD = 27.990), $t(153) = 13.229$, $p < .001$, and this involved non-zero scores on the safety voice dictionary (M = 9.474, SD = 10.348), $t(153) = 11.362$, $p < .001$. Specifically, people withholding safety concerns engaged in informative (M = 2.877, SD = 3.878), inquisitive (M = 3.714, SD =3.970), prohibitive (M = 1.75, SD = 2.771), cautionary (M = .331, SD = .724) and oblique safety voice (M = 1.617, SD = 1.931), $t(153)$s ≥ 5.680, ps < .001. The distinction between safety silence and safety voice was a matter of degree: participants that did not speak-up scored lower on the five safety voice dictionaries, $F(1,390)$s ≥ 4.900, $ps \leq .028$, η^2s ≥ .016. This illustrates safety themes are less present for safety silence and, importantly, indicates that safety silence can be scaled based on the degree of safety voice in speech.

Reducing safety silence

Supporting hypotheses 2a-c, safety silence was reduced through manipulating beliefs on safety and norms for speaking-up. However, only encouragements had a direct effect on safety voice, whilst hazard salience and responsibility modified the effect of safety concerns on safety voice. That is, concerned participants did not engage in less safety silence through manipulating hazard salience, $b = -.459$, $t(305) = -.249$, $p = .804$, and responsibility, $b = -1.061$, $t(305) = -.576$, $p = .565$. Yet encouragements reduced safety silence, $b = 4.000$, $t(305) = 2.171$, $p = .031$, with participants uttering more words on the safety voice dictionary. Underscoring the importance of assessing safety concerns, interventions only reduced safety silence for the levels of the manipulations. That is, stronger safety concerns reduced safety silence in terms of the safety voice dictionary, $b = 1.287$, $t(390) = 2.032$, $p = .043$, but only when hazards were salient, $b = 1.956$, $t(388) = 2.049$, $p = .041$, participants were discouraged, $b = 1.695$, $t(388) = 1.970$, $p = .050$, and (through a marginal effect) responsibilities were clear, $b = 1.666$,

$t(388) = 1.757$, $p = .080$. Yet, stronger concerns did not reduce safety silence when hazards were not salient, responsibilities unclear and participants were encouraged, *ns*.

Probing effects. Further analyses suggested that stronger concerns did not universally reduce safety silence: stronger concerns only reduced unique manifestations of safety silence for the levels of the manipulations. That is, salient hazards only led to more inquisitive safety voice, $b = .705$, $t(388) = 2.091$, $p = .037$; clear responsibilities to more inquisitive safety voice, $b = .698$, $t(388) = 2.083$, $p = .038$, and less oblique safety voice, $b = .226$, $t(388) = 2.157$, $p = .032$; and discouragements only to more informative safety voice, $b = .890$, $t(388) = 2.005$, $p = .046$. Otherwise, safety concerns did not reduce safety silence on the safety voice dictionaries, *ns*. This indicates that sensemaking on safety knowledge is elicited by salient hazards, clear responsibilities and encouragements, whilst clear responsibilities reduce unclear speech. Furthermore, it suggests that safety motivation-based themes are not elicited by the manipulations and may be better addressed by alternative interventions.

The effect of time on reducing safety silence

Safety silence manifested differently for participants that initially spoke-up during the first (i.e., conceptualisation stage; n = 44), second (i.e., encounter stage; n = 39) or third stage (i.e., imminent danger stage; n = 104) of the hazard, $F(1,184) = 13.686$, $p < .001$, $\eta^2 = .129$, and this indicates the need to compare the manifestation of safety silence across these stages.

Specifically, in comparison to other stages, dictionary scores during the conceptualisation stage indicated that participants were more concerned, $F(1,179) = 17.371$, $p < .001$, $\eta^2 = .086$, and this led to more informative, inquisitive and prohibitive, and less oblique and disfluent safety voice, $F(1,179)\text{s} \geq 4.846$, $ps \leq .029$, $\eta^2\text{s} \geq .026$. They did not engage in more cautionary safety voice than in the other stages, *ns*. Partially supporting **hypothesis 3**, this suggests participants at this time-point were orientated towards evaluating the idea of walking the plank, without perceived risk interrupting speech. The encounter stage only involved marginally more informative safety voice, $F(1,178) = 3.429$, $p = .066$, $\eta^2 \geq .018$. This suggests that the second stage may involve sensemaking about the physical encounter of the walking the plank idea. Finally, when danger was imminent, participants engaged in more informative safety voice, $F(1,178) = 54.728$, $p < .001$, $\eta^2 = .228$. However, their speech was also less concerned and prohibitive, $F(1,178)\text{s} \geq 8.080$, $p\text{s} \leq .005$, $\eta^2\text{s} \geq .042$. This suggests imminent harm is more effectively reduced through indicating safety knowledge rather than safety motivation. Interestingly, the imminent danger stage led to higher disfluency and oblique safety voice scores, $F(1,178)\text{s} \geq 10.704$, $p\text{s} \leq .001$, $\eta^2\text{s} \geq .055$, suggesting that mitigating imminent danger may be cognitively disruptive.

Finally, encouragements reduced the likelihood for safety silence during the first stage of the hazard, $OR = .205$, $z(305) = -2.672$, $p = .008$, whereas clear responsibilities increased the likelihood that people spoke-up during the second stage, $OR = .257$, $z(305) = 2.058$, $p = .040$.

Discussion

We proposed a model for scaling safety silence based on five types of safety voice speech and evaluated when interventions can reduce safety silence. Our experimental investigation provided the first behavioural evidence that safety silence can be scaled upon the degree of safety voice speech for concerned participants and demonstrates that interventions for reducing safety silence are most effective when participants hold safety concerns. Additionally, interventions tend to reduce knowledge-based speech, but not motivation-based speech, and that the temporal progression of hazards leads, in temporal order, to conceptual evaluations, exploration of consequences, and attempts at mitigating the hazard. These findings have implications for conceptualising and reducing safety silence.

Theoretical implications

First, we revealed that safety silence can be scaled based on the degree of safety voice speech. Specifically, safety silence manifests in the degree that concerned individuals talk about safety knowledge (i.e., informative, inquisitive) and motivation (i.e., prohibitive, cautionary), or speak unclearly (i.e., oblique speech). We outlined a model capturing these themes in speech for (un)concerned participants and evidenced that safety silence contains meaningful information (i.e., extent of concerns and in-situ inhibiting effects). This conceptual model underscores propositions and findings that safety concerns can be raised through different means (Friedman et al., 2015; Kassing, 2002; Krenz et al., 2019; Pian-Smith et al., 2009), and clarifies that the nature of safety silence is rooted in the veiling of concerns (Morison & Macleod, 2014) and degrees of thematic silence (Kurzon, 2007, 2011). Thus, we add to safety voice theory through evidencing the relationship between safety voice and safety silence in terms of five themes reflecting safety knowledge and safety motivation (Christian et al., 2009; Searle, 2008). Furthermore, through providing the first text-based measures for assessing safety silence and safety concerns, we enable a novel stream of research on silence in text. For example, in online environments (e.g., tweets containing hashtags on safety; Purohit et al., 2013), laboratory scenarios (de Ruiter & Albert, 2017; Kendrick, 2017), and field settings (e.g., cockpit voice recordings; U. Fischer & Orasanu, 2000; Sassen, 2005).

Second, we revealed that scaling safety silence is important for designing interventions. That is, through evaluating hazard salience (Tucker et al., 2008), felt responsibility (e.g., Duan et al., 2017) and encouragements (e.g., Barzallo Salazar et al., 2014; Burris, 2012), we showed that interventions are most effective at reducing safety silence when participants are concerned, and for safety knowledge-based speech. This suggests that safety voice and silence involve sensemaking and interventions need to account for participants seeking information on risk (e.g., R. J. Griffin et al., 1999): to ascertain the presence and appropriateness of discussing risks, people use in-situ information on hazards and anticipated responses. This corresponds to theory on risk communication (Lindell, 2018), harm prevention (P. Fischer et al., 2011) and safety climate (Zohar, 2010) that propose that people engage in social information processing to evaluate environmental (i.e., risk) and social cues (e.g., psychological safety; Edmondson & Lei, 2014). We expand these insights through indicating that information only elicits specific types of sensemaking and encourage future research to investigate when people are likely to discuss safety motivations.

Finally, we revealed that safety silence is time-bound. Speech reflecting safety knowledge was relevant across stages (with a marginal trend for the encounter stage). Conversely, speech reflecting motivations for avoiding harm (i.e., prohibitive safety voice) were only more prevalent when hazards were conceptual, and this may be explained by the cognitive disruption of imminent harm. This corresponds to recent findings indicating that nurses voice later, not less, dependent on leadership influences (Krenz et al., 2020) and underscores that studies need to account for the temporal nature of safety silence. For instance, research may identify how silence progresses over time in speech through directly observing how spirals of silence (Noelle-Neumann, 1974; Scheufle & Moy, 2000) emerge and change through the impact of others (e.g., through poor listening out for concerns), or applying safety voice dictionary scores to explain decision-making on uncertain risks (Brashers, 2001; Lindell & Perry, 2012).

Practical applications

Our findings indicate at least three applications. First, through providing new measures we enable the recognition of safety silence in applied settings. This is important for areas such as accident analyses, training programs, and automated speech recognition. For example, Tarnow (2000) described how the crash of Express II Airlines, Inc./Northwest Airlink 5719 was attributable to tense and hesitant communication, and Fischer and Orassanu (2000) described indirect speech contributing to the crash of Air Florida Flight 90. Because of this, training programs aimed at safety-specific communication (e.g., Crew Resource Management, teamSTEPPS, LOFT; Kanki et al., 2019; King et al., 2008) typically emphasise communication styles (e.g., assertiveness) and collaboration on safety (e.g., shared mental models, adaptability, error management; Helmreich et al., 1999). However, these rarely train on specific speech (Leonard et al., 2004), or the measurement of safety silence. We indicate that safety silence can be identified within speech, and the presented dictionaries may be applied to conversations in field settings for recognising specific speech patterns that distinguish voice and silence. Practitioners may be trained to identify these patterns (e.g., in live or recorded speech), and research should investigate whether dictionaries for laboratory-based hazards can be generalised to field settings. For instance, through developing automated speech recognition software for speech within healthcare environments (e.g., medical reports, notes; Jiang et al., 2017).

Second, we indicate that safety silence is contingent upon the perception of risk and that interventions are therefore most optimal for concerned people. This appear especially useful for altering whether people discuss safety knowledge, and silence may therefore be optimally reduced through providing explicit safety information (e.g., in healthcare leaflets, through warning signs; Matthews et al., 2014; Pander Maat & Lentz, 2010), clear accountability structures, and inclusive leadership (Nembhard & Edmondson, 2006; Weiss et al., 2018).

Finally, practitioners may utilise the experimental paradigm itself for training purposes. Debriefing simulations can raise awareness amongst participants (Kolbe et al., 2015) and corresponding to this, we noted that the debriefing appeared to increase knowledge on the impact of beliefs on safety and perceived norms for speaking-up. The scenario may be especially relevant for training on mild risks and for people with no specific safety knowledge (e.g., patient, customers): the scenario does not require specialist knowledge and has a low threshold for participation.

Limitations

First, the experimental paradigm has debated external validity (Noort et al., 2019a). A need remains to extend findings to natural speech in other contexts (e.g., in operating rooms, flight decks) and to scenarios that pose more substantial risk. This appears key because speech is highly context-dependent (e.g., informative speech included characteristics of the experimental scenario; Gillespie & Cornish, 2010). Yet, the criteria to establish fidelity are not clear (Nestel et al., 2017) and the standardised scenario reveals causal mechanisms with high internal validity that may be generalised with more certainty to settings with known characteristics (Noort et al., 2019a). We argue that the method is appropriate because hazards are presented ethically (through the perception of risk) and elicited safety concerns. To evidence the generalisability of concepts, we aim to establish safety voice during real-world hazards in future research.

Second, the responsibility manipulation unexpectedly reduced felt obligation. This indicates interpretation of this manipulation is not straightforward and future research should examine this. For instance, reminding people of previously held responsibilities may have compensated the need to feel responsible in novel situations. Still, the manipulation reduced safety silence and led to more inclusive language indicating that participants may have felt shared rather than individual responsibility.

Finally, the association between concerned speech and safety voice was very strong and this indicates that safety concerns and safety voice, though distinct concepts, may be less distinguishable in speech. Future research may expand on this.

Conclusions

Reducing safety silence is critical for safety management. We proposed a model for the manifestation of safety silence behaviour and showed that safety silence is scalable in terms of the degree of safety voice speech. Furthermore, we showed that interventions only reduced safety silence about safety knowledge when participants were concerned, and at specific time-points. Our findings underscore the importance of the behavioural investigation of safety silence and the need to assess the extent people perceive risks: safety silence is reduced most effectively when safety information is available, and this is manifested in speech. Future accidents may therefore be prevented through investigating the manifestation of safety silence in speech.

CHAPTER 6:

APPLYING THE CONCEPT OF SAFETY VOICE TO REAL-LIFE HAZARDS: THE NEED FOR SAFETY LISTENING

Preface

In the previous chapters, I built a behavioural methodology and a conceptualisation of safety voice as a nuanced behaviour that is deeply rooted within characteristics of the hazardous situations in terms of how safety threats elicit safety concerns and safety voice. For instance, I indicated that the behavioural nature of safety voice is characterised by ecological relationships to other variables (e.g., hazard-related antecedents; Chapter 3), and through a simulated hazard I revealed that safety voice is contingent on the extent to which people perceive risk within a hazardous situation (Chapter 4), the extent to which situational characteristics elicit variation in talk about safety, and the temporal progression of hazardous situations (see Chapter 4 and 5). These findings contribute to safety management by underscoring that accidents can occur when safety voice is inhibited by social relationships (e.g., when there is low psychological safety; Edmondson, 1999) and the characteristics of hazardous situations (e.g., when hazards are not perceived as risky), with interventions needing to be tailored accordingly (e.g., to safety concerns and safety voice).

However, a need exists for establishing and explaining the extent to which safety voice manifests during real-life, actual, hazards, and how this enables the mitigation of accidents when others listen. In the previous studies I did not present actual hazards but observed voice behaviour in simulated environments, and the literature has not clarified to what extent the concept of safety voice may be applied to real-life hazards (Krenz et al., 2020; Schwappach & Gehring, 2014c). For instance, using the Walking the Plank paradigm (n = 55), MSc students at LSE[27] were able to demonstrate that closed leadership inhibits safety voice behaviour, $OR = .145$, $Wald(1) = 9.435$, $p = .002$. However, similar to the studies in previous chapters and research by Barzallo Salazar and colleagues (2014), risk was operationalised in terms of perceived risk and simulated hazards, not actual hazards, and it controlled how others listened to safety voice. This practice ensures ethical standards (see Chapters 2 and 4), yet because of this, it remains assumed that the concept of safety voice is important for understanding how accidents are caused. Furthermore, it is unclear when safety voice is effective for preventing harm (e.g., when others listen), and even if safety voice occurs at all during actual hazards. This is important, because if safety voice may be understood as highly situated and nuanced behaviour important for managing safety threats, it should emerge in response to extreme levels of risk (e.g., fatal) posed by real-world accidents and in the context of others actually listening or dismissing safety concerns (‘safety listening’).

Thus, in this chapter I set out to investigate the extent to which the concept of safety voice is important for understanding accident causation, and I present a study using naturally occurring data from historic aviation accidents. Aviation accidents provide naturally occurring cases where actual risk was confirmed to be present and high, with Cockpit Voice Recorder (CVR, also known as the ‘black box’) transcripts providing field-based data on communication prior to aviation accidents whereby participants do not need to be exposed to new risks. In addition, the global and hierarchical nature of the aviation industry enables the investigation of the extent to which safety behaviours during accidents are shaped by social characteristics such as norms for communicating with seniors (e.g., power distance). Finally, because included accidents occurred between 1962 and 2018, I was able to investigate the extent to which safety voice and safety listening changed during a historic period when safety management improved

[27] This involved: Eleanor McSweeney, Grace Rahal, Kate Sitniewski and Lucy Zhong. I independently analysed the presented statistic.

significantly through the introduction of new theoretical models outlining how sociotechnical systems can be designed to prevent harm (Perrow, 2011; Reason, 2000; Vaughan, 1996; Zohar, 1980), and training programs that aimed to improve how teams coordinate on safety (e.g., Crew Resource Management; Kanki et al., 2019).

Thus, in this chapter, I advance safety management by clarifying how safety voice is relevant for understanding accident causation, evaluating the extent to which others listen to safety concerns, and providing insights into leadership effects on voice during accidents (Detert & Treviño, 2010). This enables the evaluation of the extent to which safety voice is highly situated, and a conceptualisation of accident causation that specifies that accidents may occur when hazards do not elicit an effective chain of safety concerns, safety voice and safety listening (therefore contributing to research question 1: *What is the behavioural nature of safety voice?*). Furthermore, in this chapter I illustrate how safety voice may be investigated in the field by utilising archival data and indicate the extent to which findings on safety voice and leadership can be generalised to genuine accidents (therefore contributing to research question 2: *What is the optimal way to investigate safety voice behaviour?*). Finally, because interventions that aim to improve coordination and decision-making on safety such as CRM (Kanki et al., 2019) have been introduced across the aviation industry from the 1980s (Helmreich et al., 1990), the historic and global nature of the data enables an evaluation of the effectiveness of CRM training for improving safety voice and safety listening across contexts that vary in power distance (therefore contributing to research question 3: *To what extent do interventions for promoting safety voice and reducing safety silence need to be tailored to the behavioural nature of safety voice?*).

Study description

Article 4 presents a study of safety voice during historic aviation accidents. A dataset was generated from 172 CVR transcripts that were available in three online databases. Transcripts were included in accident investigation reports and provided direct access to communication during actual accidents (n = 14,128 spoken messages). By evaluating the extent to which flight crew spoke-up across and during accidents and triangulating this data with safety listening and Hofstede's Power Distance Index (e.g., Hofstede et al., 2010), I indicated that flight crew nearly always spoke-up across accidents (99%), and this was significantly higher compared to levels of safety voice established in the Walking the plank paradigm (see Chapter 4) and the systematic literature review (i.e., both 44%; see Chapter 3). This indicated that a lack of safety voice cannot be assumed to have caused accidents or be essential for preventing accidents, and it underscored the need to understand how safety voice can be made more effective. That is, accidents still occurred in the presence of safety voice, and findings underscored that this should emphasise safety listening because poor safety listening reduced the engagement of junior flight crew in safety voice. This provides the first evidence of the proposition by Barlow and colleagues (2019) on the essential role of 'receivership' (which I term 'safety listening')

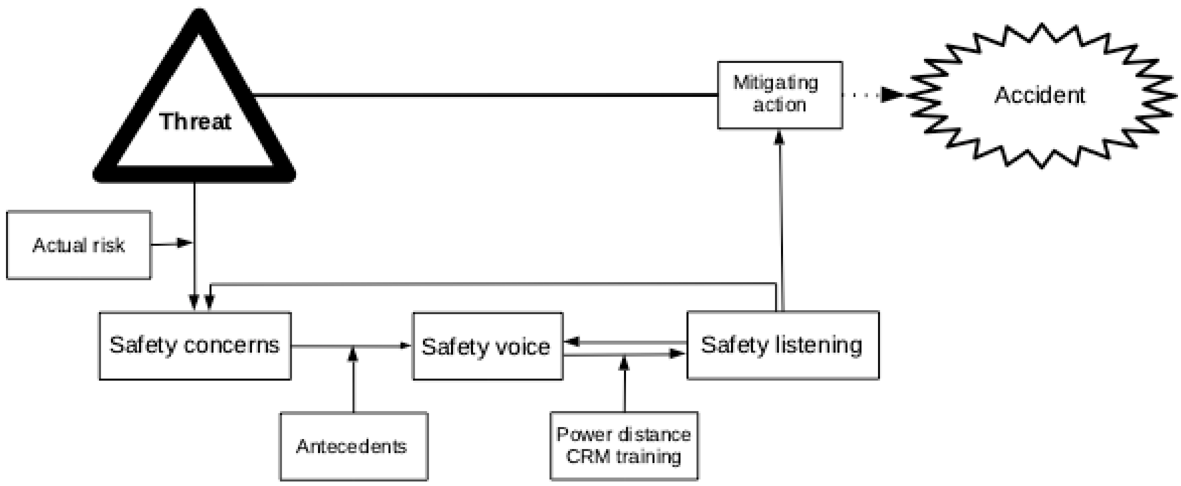


Figure 6.1. Threat Mitigation Model of safety concerns, safety voice and safety listening.
Note: this model highlights that the dysfunctional momentum of threats towards accidents (Barton & Sutcliff, 2009) can be mitigated when safety threats successfully elicit safety concerns, safety voice and safety listening, with antecedents impacting on the relationships.

in mitigating accidents (this also enables the investiagation of leadership concepts from the field of employee voice within the context of real accidents; Detert & Treviño, 2010). Moreover, it contributes to safety management by extending the safety voice model developed across the previous chapters: 'safety listening' is an important step in mitigating safety threats through interpersonal communication during hazardous scenarios (see Figure 6.1, and Figure 6.4 which presents the conceptual model without specified antecedents).

Moreover, junior flight crew varied in the extent to which they spoke-up during accidents (e.g., once versus repeated), and this underscored the need to understand the role of safety voice for safety management as highly situated: junior flight crew spoke-up more in airlines from low power distance countries (i.e., indicating the importance of national culture) and when safety listening was stronger (i.e., indicating the need for safety listening interventions).

Again, underscoring the situated nature of safety voice, results indicated that the extent of safety voice (during accidents) became less over the years, and this was explained by safety listening improving over time where power distance was low. This is important for safety management because it suggests that communication about safety concerns became more effective (i.e., less safety voice was necessary) and may even have prevented accidents in low power distance countries (i.e., because voice occurs more for near-misses than accidents; Blanco et al., 2009). Furthermore, it indicates that CRM training programs have only been partially effective: safety voice and safety listening changed after the introduction of these training programs in the early 1980s, but only for low power distance countries.

Authorship

I was responsible for designing the study, preparing the coding framework, analysing and interpreting the data and preparing the manuscript (80%). Dr Tom Reader and Dr Alex Gillespie contributed to conceptualising the study, the design of the coding framework and manuscript preparation (20%). Research assistants (Lindsie Arthur-Hulme, Alex Goddard, Nanne Houtsma, Celestin Okoroji) collected and coded the data and contributed towards the coding for the interrater reliability analysis.

Article 4:
Safety voice and safety listening during aviation accidents: when speaking-up to power is not enough

Submitted for publication in:

Risk Analysis: an international journal (submission date: 22 July 2020)

Abstract

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

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

Introduction

Safety voice is the act of speaking-up about perceived hazards (Noort et al., 2019b; Tucker et al., 2008). For high reliability industries such as aviation, safety voice is assumed to be central to maintaining safe operations (Bienefeld & Grote, 2012) and where team members withhold safety concerns (‘safety silence’), or dismiss them (i.e., poor ‘safety listening’), this has contributed to tragic accidents due to information about risk not being shared or used (Cocklin, 2004; Cookson, 2015; NTSB, 1978). Explanations for the absence of safety voice and poor safety listening during safety critical scenarios often focus on cultural norms and asymmetric leader-follower relationships (i.e., power distance; Hofstede et al., 2010). Specifically, accidents are assumed to emerge from people not speaking-up due to fears for the social consequences of incorrectly raising concerns or undermining leaders (Enomoto & Geisler, 2017; Gladwell, 2008; Soeters & Boer, 2000; Tucker et al., 2008), and poor safety listening to voice is understood to arise from norms for communication (Hofstede et al., 2010; Kam & Bond, 2009) and expected asymmetries on expertise for managing safety (Tost et al., 2012). Studies utilising vignette (Schwappach & Gehring, 2014c), laboratory (Noort et al., 2019a), high-fidelity simulator scenarios (Barzallo Salazar et al., 2014) and case studies (Driscoll, 2002) have explored this extensively, and show that power dynamics shape how leaders respond to advice (Tost et al., 2012), and that when leaders listen poorly to safety concerns (Edmondson, 1999; Nembhard & Edmondson, 2006), junior team members are less likely to engage in safety voice, or delay speaking-up (Krenz et al., 2020), which impairs safety management.

Thus, safety voice and power distance are recognised as primary causes of organisational accidents (Conchie et al., 2012; Enomoto & Geisler, 2017; Gladwell, 2008; Reader et al., 2015; Soeters & Boer, 2000), and a range of interventions for reducing power distance in teams and enhancing speaking-up (e.g., psychological safety, training; Kanki et al., 2019; Kolbe, Weiss, et al., 2013; Newman et al., 2017) have been developed to improve safety voice and safety listening. Although laudable in intention, this remains grounded in little empirical evidence for the role of safety voice as a causal factor in accidents, and the influence of power distance upon the behaviour. Safety voice has largely been investigated through case studies, experiments, and self-report data (Noort et al., 2019a), and research has not established the extent to which an absence of safety voice, or poor safety listening, have directly contributed to accidents where actors (e.g., flight crews, patients) experienced serious threats to life (e.g., Raemer et al., 2016) outside of isolated accident investigations (e.g., Francis, 2013; NTSB, 1978). Determining this is important for ensuring that theory on the contribution of safety voice and power distance to accident causation, and interventions that flow from this, are grounded in observations examining how and to what extent they contribute to real accidents.

We address this evidence gap in the current study, and through analysing cockpit voice recorder (CVR) transcripts of 172 historic aviation accidents, examine the role and nature of safety voice behaviours in accident mitigation. We establish to what extent safety voice i) manifests prior to accidents, ii) is ignored or dismissed by crew members, and iii) is explained by cultural norms for how junior and senior crew interact (i.e., power distance). We also consider how the introduction of Crew Resource Management (Helmreich et al., 1990; Kanki et al., 2019), an intervention designed to improve teamwork amongst safety-critical staff (e.g., flight crews, critical care teams), has increased safety voice. Our contribution is to systematically establish the role of safety voice, safety listening and power distance in real accidents, and through this, advance understanding on to extent to which a lack of ineffective communication contributes to accidents.

Safety voice for safety-critical staff

Safety voice is the act of speaking-up about perceived hazards to others of equal or senior status in order to mitigate harm (Noort et al., 2019b; Tucker et al., 2008). Conversely, when people withhold safety concerns this is labelled 'safety silence' (Manapragada & Bruk-Lee, 2016). The concept draws from research on communication and safety management (Noort et al., 2019b) and especially employee voice research (Morrison, 2014; Noort et al., 2019b; Tucker et al., 2008). The central idea to this research is that individual team members may have critical information (e.g., on risk), and that the free flow of this information contributes to mitigating failures (Westrum, 2014). Because of this, and the harmful consequences of poorly sharing safety information (e.g., Kolbe et al., 2012; Novak, 2019), scholars have distinguished the concept of safety voice and provided a distinct literature (Conchie et al., 2012; Morrow et al., 2016; Noort et al., 2019b; Okuyama et al., 2014; Tucker et al., 2008) that extends beyond organisational environments (e.g., to non-smokers in public settings; Bigman et al., 2019), provides unique empirical data (Noort et al., 2019a), and captures the communication of safety concerns that emerge from perceived risks (e.g., Schwappach & Gehring, 2014c).

Safety voice is of vital importance to environments where people need to decide and act on perceived risks, such as flight crews, nuclear control room teams, critical care teams, or oil rig maintenance teams. Highlighting unsafe conditions helps to interpret the environment, create shared situational awareness (Driscoll, 2002; Foushee, 1984), enables mitigating actions (Barton & Sutcliffe, 2009; Sexton & Helmreich, 2003), and improves safety performance (Manapragada & Bruk-Lee, 2016; Manias, 2015), especially when junior members of technical teams speak-up (Kolbe et al., 2012). For instance, in aviation, flight crews continuously handle routine and novel hazardous scenarios (e.g., taking off in poor weather, deciding on warning signals), and voicing and listening to concerns is deemed necessary for avoiding potentially fatal outcomes (e.g., crashes, mid-air collision). That is, operating aircraft requires effective coordination (e.g., to decide on risk, complete checklists, avoid opposing system input, etc.; Skogstad & Hellesoy, 1995) between pilots that share responsibilities for maintaining safe flight, yet have distinct tasks (e.g., flying the aircraft, monitoring radio communication), information (e.g., duplicated meters may provide divergent information), experience and seniority.

Ineffective crew coordination, though rarely the sole causal factor, has contributed to accidents through loss of situational awareness and ineffective decision-making. For instance, status differences and poor coordination (e.g., the lack of raising or listening to safety concerns) contributed to fatal accidents in healthcare (e.g., the death of Elaine Bromiley after concerns about a difficult airway were dismissed; Bromiley & Mitchell, 2009; Fioratou et al., 2010), aviation (e.g., the crash of United Airlines 173 after fuel starvation was ignored; NTSB, 1978) and energy (e.g., the blow out of the Deepwater Horizon oil rig after concerns about a pressure test were not raised by contractors; Reader & O'Connor, 2014). Thus, the role of communication problems in accidents within a broad range of domains underlie the growth of the safety voice literature, and the focus of interventions to reduce safety silence and poor safety listening.

To explain why junior team members do not engage in voice, and why senior team members do not listen effectively, studies have drawn on the concept of power distance (Botero & Van Dyne, 2009; Gladwell, 2008; Huang et al., 2005; Kwon et al., 2016; Landau, 2009; Liang et al., 2012; MacNab et al., 2007; Morrison, 2011, 2014; Newman et al., 2017; Rhee et al., 2014; Wilkinson et al., 2020), which "refers to the degree to which individuals, groups, or societies accept inequalities (…) as unavoidable, legitimate, or functional" (Daniels & Greguras, 2014, p.2). Studies indicate unfavourable effects of power distance for communicating issues to leaders (Peadon et al., 2020) and subsequent interventions aim to enable leaders to listen better to safety voice (e.g., support, enacting change). Yet, researching this is challenging because safety voice emerges spontaneously and its infrequent occurrence cannot be readily controlled (e.g., prompting voice could bias findings; Noort et al., 2019a). To address this, and because introducing real hazards is unethical (APA, 2017), research has assessed safety voice through hazardous scenarios in interviews, focus-groups and surveys (e.g., prompting memories; Alingh et al., 2019; Manapragada & Bruk-Lee, 2016), vignettes (Schwappach & Gehring, 2014c), high-fidelity simulations (e.g., during technical procedures; Barzallo Salazar et al., 2014; Foushee & Manos, 1981; H. P. R. Smith, 1979; Weiss et al., 2018), simulation-based training (Kanki et al., 2019; Kines et al., 2010; Kolbe, Weiss, et al., 2013; Leonard et al., 2004) and through laboratory experiments (e.g., presenting generic risks; Noort et al., 2019a). These approaches generated the insight that safety voice can be promoted (in terms of likelihood or onset) through leaders acting in low power distance ways. For instance, through providing encouragements, using inclusive language (Barzallo Salazar et al., 2014; Weiss et al., 2018) or shallower hierarchies (Krenz et al., 2020). Furthermore, the perception of risk may be necessary for successful interventions (Noort et al., 2019a), and the decision to raise safety concerns is a trade-off between the benefit of mitigating harm and the cost of leaders' poor listening to safety voice (Schwappach & Gehring, 2014c).

Yet, insights on the extent to which, and how precisely, safety voice contributes to real accidents remain scare and limited (Krenz et al., 2020; Noort et al., 2019a; Peadon et al., 2020), and this is especially true for the degree that leadership shapes safety voice during naturally occurring scenarios. Insights are largely drawn from case studies that are selectively chosen because of the role of safety voice in accident causation (e.g., Driscoll, 2002), studies that operationalise accidents instead of safety voice behaviour (Anicich et al., 2015; Enomoto & Geisler, 2017; Soeters & Boer, 2000), or inquests (Francis, 2013, 2015) that may poorly reveal actual safety voice behaviour because data reflects participants' perspectives on historic events (Noort et al., 2019a). Available evidence may therefore not accurately represent voice during accidents, and the problem posed by safety voice may be overestimated (e.g., if the frequency of safety voice is biased). Subtleties, like the strategies used to voice safety concerns and the ways in which voice is dismissed, have, to our knowledge, never been investigated in the context of real accidents. This has led to the widespread assumption that safety voice is a substantial contributory factor to accidents, and is therefore important for mitigating declining conditions, errors and accidents (e.g., employee voice, safety voice, psychological safety; Edmondson & Lei, 2014; Morrison, 2014; Okuyama et al., 2014), and a function of wider organisational environments (e.g., safety culture, safety citizenship; Reader et al., 2015; van Dyne et al., 2003). However, to date, there is no systematic exploration of the extent to which a lack of safety voice and poor listening (i.e., failures to report safety problems, dismissal of concerns) are a central and major contributing factor to serious accidents (Krenz et al., 2020), and the level of influence exerted by power on safety voice (rather than, for example, time, failure to notice information, error) remains a proposition (e.g., Kwon et al., 2016).

Thus, whilst safety voice theory aims to explain how the behaviour contributes to accidents, and to develop interventions for improving speaking-up (O'Donovan & McAuliffe, 2020), there is a lack of data on to the extent to which, and how precisely, safety voice manifests and is listened to during accidents. Given the importance of safety voice as a conceptual frame for explaining failures in safety management, and for training programs aiming to improve coordination on safety (e.g., crew resource management, TeamSTEPPS; Kanki et al., 2019; King et al., 2008), it appears essential to consider its actual role in accident causation. For instance, without this, it is unclear how field-based behaviour should be mapped unto survey findings, or to what extent findings on power distance and safety listening generalise to the field. This idea is consistent with the broader observation from meta analyses that, whilst flexible approaches to methodological realism are appropriate (i.e., there are no strict rules for what makes scenarios "real"; Nestel et al., 2017), psychological effects established in controlled settings (e.g., simulation or laboratory studies) can be substantially different in the field and vary in their direction (Mitchell, 2012), with effects on safety-related behaviour being stronger in archival data than self-reports (Christian et al., 2009). Specifically, and as we describe below, evidence is lacking for accidents where teams were not able to manage real and severe risk in terms of the extent to which people engage in safety voice and safety listening, or how power distance may explain this. We explore this through analysing CVR data capturing speech acts on flight decks prior to accidents.

The current study

Here, we investigate the extent to which safety voice varies during actual hazards that pose extreme risk, and how safety listening and power distance shape this. For aviation, this may be achieved through analysing transcripts from cockpit voice recorders (CVRs) from historic aviation accidents. CVRs were designed to capture and interpret sounds during accidents (e.g., flight crew communication, cues on hazards; Maher, 2018), and research on flight crew communication (U. Fischer & Orasanu, 2000; Foushee & Manos, 1981; Nevile & Walker, 2005; Orasanu & U. Fischer, 1992; Sassen, 2005; Sexton & Helmreich, 2000) indicates CVR transcripts can be used to analyse in-situ interactions between flight crew. Thus, utilising CVR data, we investigate the degree to which aviation accidents elicit safety voice, and address three outstanding issues.

Safety voice during aviation accidents

Safety voice occurs in the context of hazards, and the mitigation of risk through speaking-up is central to the concept of safety voice. Typically, hazards are operationalised through actual risk being hypothesised (e.g., for vignettes, simulations; Krenz et al., 2020; Schwappach & Gehring, 2014c) or controlled (e.g., for laboratory scenarios; Noort et al., 2019a). This revealed that stronger risk perceptions (e.g., for missed hand disinfection) are associated with more safety voice (Gurung et al., 2017; Schwappach & Gehring, 2014a, 2014c). Yet, through presenting scenarios with minimal risk, perceived risks are assumed to generalise to risk during actual hazards. Because this remains undetermined, however (Krenz et al., 2020), we do not know the extent to which visceral responses to risk (e.g., dread, fear) elicit safety voice (Loewenstein et al., 2001; Noort et al., 2019a). Establishing this is important because behavioural variations can indicate when intervention may be successful (e.g., if power distance shapes safety voice). Conversely, very frequent or infrequent safety voice would suggest, respectively, that the behaviour is ineffectual and interventions should improve safety voice's effectiveness (e.g., when recipients listen), or that speaking-up does not pose a problem for accident causation (e.g., because it always mitigates harm, or risk simply does not elicit safety voice in practice).

We propose that actual hazards, and especially fatal accidents, lead to more safety voice than typically established in the literature (i.e., approximately 44% of concerns are raised; Noort et al., 2019b) because cognitive evaluations of risk and visceral affective states of dread motivate stronger behavioural responses to mitigate harm. Probabilistic risk models highlight that hazards emerge from the accumulation of sociotechnical factors (e.g., resourcing, systems design, unsafe acts; Leveson, 2011; Reason, 2000), with greater risks (i.e., impact and likelihood; Renn, 1992) increasing the need for mitigating action. Yet, technical properties of risk are often difficult to evaluate (e.g., because information is ambiguous; Viscusi & Zeckhauser, 2015) and the psychometric approach therefore highlights that the response to hazards is rooted in analytic and affective risk perceptions (Slovic, 1987; Slovic et al., 2004). Visceral affective states emerge where encountered risks are fatal, involuntary and personally relevant, with affect heuristics providing a strong motivation to alter unsafe conditions (Loewenstein et al., 2001; Slovic, 1987, 2016). This is important, because safety voice theory often explains behaviour in terms of employee motivation (e.g., safety participation or safety citizenship; Christian et al., 2009), and little analysis has considered motivations that emerge from potentially fatal contexts. Furthermore, high costs of speaking-up may be rationally traded-off with the larger cost posed by fatalities (Noort et al., 2019b; Schwappach & Gehring, 2014a) because the higher expected utility of speaking-up increases voice (Murphy & Dingwall, 2007b). Thus, and in contrast to the literature's assumption that accidents emerge from relatively low levels of safety voice (Bienefeld & Grote, 2012; Enomoto & Geisler, 2017; Gladwell, 2008; Kolbe et al., 2012; Noort et al., 2019b; Soeters & Boer, 2000; Tucker et al., 2008), flight crew may frequently engage in safety voice due to the extreme level of risk posed by aviation accidents. To investigate this, we ask: *1a) to what extent did flight crew engage in safety voice across historic aviation accidents?*

Furthermore, we examine whether flight crew engagement in safety voice prior to accidents has changed over the course of history. Within the safety literature, the training of interpersonal skills is widely seen as key for improving safety voice and safety-related attitudes (O'Connor et al., 2008), and in aviation such training has been in place since the early 1980s through the implementation of Crew Resource Management training programs ("CRM"; Helmreich et al., 1999; Kanki et al., 2019). Over time, these training programs became widespread (O'Connor et al., 2008, 2012) and increased in effectiveness through emphasising the design of social environments (e.g., teamworking and organisational culture) in addition to the correction of human error (Helmreich et al., 1999). CRM implementation may therefore have increased flight crew engagement in safety voice, and establishing this within the CVR data may inform the effectiveness of interventions for increasing safety voice. Thus, we ask *1b) to what extent has safety voice increased during aviation accidents since the 1980s?*

Poor safety listening

Because safety voice is aimed at others of equal or senior status, the field has aimed to identify when leadership practices are favourable for speaking-up (Detert & Treviño, 2010). Ample research indicates that the extent to which seniors listen effectively to safety voice (e.g., acknowledging and acting on concerns, versus ignoring or dismissing concerns) promotes subsequent voice (Edmondson, 1999; Nembhard & Edmondson, 2006). For instance, junior staff are especially more likely to speak-up (Nembhard & Edmondson, 2006), or to do this sooner (Krenz et al., 2020), when leaders are expected to listen (Edmondson, 1999; Newman et al., 2017) and indicate that speaking-up is appropriate through acting in inclusive and encouraging ways (Barzallo Salazar et al., 2014; Bienefeld & Grote, 2014; Weiss et al., 2018). However, seniors tend to poorly listen to advice from junior staff (e.g., due to the social cost of advice-taking; Tost et al., 2012). This suggests that even if safety voice occurs frequently it may not be listened to, with poor safety listening (i.e., ignoring or dismissing safety concerns) emerging when concerns are deemed inappropriate (e.g., when concerns are considered as factually incorrect or violating social norms; Kam & Bond, 2009). For instance, no relationship between safety voice and safety listening would indicate safety voice is better predicted by risk perceptions than acting in socially appropriate ways. Conversely, when poor safety listening reduces safety voice during accidents this would indicate that risk perceptions only partly explain safety voice and that social motivations shape the behaviour, even during extreme personal risk. If so, unique interventions are required for safety listening as a distinct contributor to accidents, and safety voice behaviour would be central to situated sense-making on risk: people share and decide on perceptions about encountered hazards, with voicing and listening to safety concerns providing two distinct aspects of a larger phenomenon capturing on-going, dynamic safety conversations. Evaluating safety listening is therefore important for conceptualising safety voice, and thus, we ask: *2a) to what extent does safety listening predict safety voice engagement for junior staff during aviation accidents?*

Furthermore, we also examine whether safety listening has changed over the course of history for flight crew. CRM training goals include improving how leaders engage in effective coordination on safety information inside the cockpit (Kanki et al., 2019). Thus, because CRM became more widespread and effective (Helmreich et al., 1999; O'Connor et al., 2008, 2012), it may be expected that safety listening improved, and establishing this is important for enabling interventions that make safety voice more effective. Thus, we ask 2b*) to what extent has safety listening improved during aviation accidents since the 1980s?*

The role of power distance

Safety voice occurs in hazardous situations that provide technical and social factors contributing to risk (Reason, 1990; Wiegmann & Shappell, 2016), and safety voice may be shaped by norms that outline how juniors communicate concerns to seniors. Ample research indicates that egalitarian relationships between leaders and followers promote open communication, and whilst the operationalisation of culture through dimensions such as power distance is debated (e.g., dimensions underrepresent cultural heterogeneity; Hofstede, 2002; McSweeney, 2002), power distance has provided fruitful hypotheses to explain variation in indicators of safety performance such as accident rates (Enomoto & Geisler, 2017; Soeters & Boer, 2000), fatalities (Anicich et al., 2015) and safety culture (Reader et al., 2015).

Furthermore, power distance has been considered in relationship to voice (Kwon et al., 2016; Liang et al., 2012; Morrison, 2011, 2014; Wilkinson et al., 2020). For example, flat hierarchies (Frosch et al., 2012; e.g., Malloy et al., 2009; Noland & Carmack, 2015) and a constructive 'tone at the top' (M. S. Schwartz et al., 2005) promote safety voice, and evidence indicates that employee's power distance orientation (an individual-level construct) reduces voice (Botero & Van Dyne, 2009; Huang et al., 2005; Landau, 2009; Rhee et al., 2014). Hofstede's power distance (Hofstede et al., 2010) may therefore provide a valuable proxy for investigating both safety voice and safety listening on the flight deck. Yet, little behavioural evidence exists, especially during actual hazards, because research on the individual-level metric of power distance orientation (Botero & Van Dyne, 2009; Huang et al., 2005; Landau, 2009; Rhee et al., 2014) has only used survey-based data for recalled and anticipated voice (i.e. preventing conclusions on behaviour; Noort et al., 2019a) and nation-level studies on power distance have operationalised accidents instead of safety voice behaviours (Enomoto & Geisler, 2017).

The power distance proposition for accident causation suggests that power distance explains accidents rates (Enomoto & Geisler, 2017) because strong norms dictate deference to seniors' authority on safety issues (Botero & Van Dyne, 2009), which are ultimately their accountability (Tucker et al., 2008). This reduces safety voice for junior flight crew (Gladwell, 2008) through high power distance "(i) discouraging the correction of errors by superiors, (ii) placing primacy of communication and debate on a superior, (iii) generating unwillingness to challenge authority, and (iv) creating asymmetrical communication between management and subordinates" (Reader et al., 2015; p.775). Additionally, safety listening may explain the relationship between power distance and safety voice because violating social norms can elicit anger (Kam & Bond, 2009) and the social cost for taking advice (e.g., appearing incompetent; Tost et al., 2012) may be higher and elicit stronger responses to juniors speaking-up where power distance is higher. However, in the absence of direct evidence, we currently do not know the role of power distance for safety voice and safety listening during critical incidents. Thus, here we examine whether wider social norms on power distance shaped behaviour in the cockpit, and ask: *3) to what degree does power distance explain safety voice and safety listening during historic aviation accidents?*

Method

Dataset

A new dataset was generated from transcripts available in published air crash investigation reports. By January 2018, 372 transcripts were obtained from three online databases (Aviation-Safety Network, 2019b; Plane Crash Info, 2019; Tailstrike, 2019). After removing duplicate, irretrievable and non-English transcripts, the final dataset contained 172 transcripts, with a total length of 21,626 lines of transcript.

Data was extracted from included transcripts: i) flight number, ii) date of incident, iii) audio source, iv) airline country registration, v) incident airspace, vi) flight phase, vii) crew and passenger numbers, viii) fatalities, ix) damage, x) attributed causal factors, xi) transcript conversational turn, xii) speaker. To provide interpretative context, narrative summaries and legends were included. In addition, each transcript line was coded using transcript legends and a coding scheme in terms of: i) turn number (i.e., sequential within transcripts), ii) turn type (i.e., conversation, background sounds, notes/information), iii) conversational turn (i.e., sequential for conversation turns within transcripts), iv) person speaking (captain, first officer, flight engineer, flight crew with unclear role, cabin crew, air traffic control, other aircraft, ground operations, other), v) the hazard raised (i.e., if one was raised, using the words of the conversational turn), vi) how others listened to the hazard raised (action, affirmed, disaffirmed, ignored, unclear), and vii) the type of hazard based on air traffic control classification schemes (i.e., ATC interaction, Crew interaction, Distraction, Equipment/fuel, Location, Manoeuvring, Weather, Pilot actions, Planning, Company actions, Other/unclear; NATS, 2013)[28].

---

[28] The NATS causal factor scheme is specific to aviation incidents but may map unto typologies with a broader application. For instance, unto levels 1-3 of the Human Factors Analysis and Classification System (Shappell & Wiegmann, 2000): 1) unsafe acts (*Manoeuvring, Pilot actions*), 2a) unsafe environmental preconditions (*Weather, Location, Equipment/fuel*), 2b) unsafe operator preconditions

Table 6.1. Attributed causes of included accidents.

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

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

Accidents in the dataset occurred between 1962 and 2018 with 97% of the cases leading to substantial damage or the destruction of aircraft, and fatalities totalling 11,001. A crude estimation puts this at approximately 15% of historical aviation fatalities in commercial and corporate aviation since 1962[29]. Most accidents occurred on approach (32.0%) or *en route* (32.0%) and were attributed to pilot actions (32.6%), see Table 6.1. Flights had an average crew of 7.120 (SD = 5.182) and 89.701 passengers (SD = 97.018), with on average 42.095 survivors (SD = 90.191). Included flights were from airlines registered in 42 countries with an average power distance of 49.103 (SD = 17.043; *range*: 11-104; *skewness* = 1.157, SE = .194).

Transcript text was based on audio sources from Cockpit Voice Recorders and/or Air Traffic Control radio communication and existed of verbal conversational turns (n = 19,393, m = 112.750; SD = 124.829) and other data (n = 2213; m = 12.866; SD = 14.452; e.g.,

(*distraction*), 2c) unsafe personnel preconditions (*ATC interaction, Crew Interaction*) and 3) unsafe supervision (*Company actions, Planning*).

[29] Aviation-Safety Network lists 66,682 historical fatalities in commercial and corporate flights between 1962-2018 (Aviation-Safety Network, 2019a), yet the full number of aviation fatalities is uncertain.

Table 6.2. Frequencies of role for speakers of conversational turns.

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

background sounds, transcriber notes). Flight crews (i.e., captains, first officers, flight engineers) provided 74.3% of the conversational turns (see Table 6.2). For the current study, analyses were performed on conversational turns from flight crew with an identified role (i.e., conversational turns from captains, first officers, flight engineers; n = 14,128), with transcripts averaging 106.001 conversational turns (*SD* = 51.727, range: 1-641). Four transcripts had less than 5 conversational turns. The full and coded dataset is provided as supplemental material.

Measures

Safety voice. Research assistants were trained on recognising safety voice through discussing illustrative examples and problematic cases, and the application of the coding scheme. They coded whether each conversational turn contained a hazard and described the hazard. If a turn contained a hazard this was coded as *safety voice* (1), otherwise this was coded as *not safety voice* (0) instead of 'safety silence' (i.e., this requires data on the extent to which flight crew were concerned). Illustrative examples are provided in Table 6.3. Good interrater reliability for safety voice was indicated for two randomly selected transcripts providing 291 conversational turns (*Gwet AC1* = .62, 95CI: .53-.71).

Table 6.3. Illustrative extracts from CVR transcripts for safety voice and response to safety voice.

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

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

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

Power Distance. Power distance was operationalised through Hofstede's Power Distance Index (PDI; Hofstede et al., 2010). PDI scores from 2015 (Hofstede, 2015) were obtained for airlines' country registration where available, bar a United Nations flight.

Safety listening. For every conversational turn containing safety voice, research assistants coded how others responded within the following three conversational turns (for illustrative examples, see Table 6.2). If a response to safety voice remained absent it was coded as ignored (0), if others disagreed or responded negatively it was coded as disaffirmed (-1), and favourable responses were coded as verbally affirmed (1) or immediate action (2). Indicating construct validity, poor safety listening was associated with accident investigation reports attributing the accident to poor crew communication (Spearman $r = -.156$, $p = .050$).

Results

Safety voice during aviation accidents

Flight crew engaged in safety voice across accidents, but the degree of safety voice was low within transcripts. Safety voice occurred in all but two of the accidents (95CI: 97.2-100.5%), with only two accidents having no instances of safety voice (i.e., Air India 182, TAM 3054). This was not different from 100% ($t(170) = -1.418$, $p = .158$). Safety voice comprised 9.78% of the conversational turns (95CI: 9.29-10.27%; $t(14085) = 390.065$, $p < .001$) and was skewed towards no safety voice (skewness = 2.709, SE = .021). The proportion safety voice for flights where someone spoke-up, and that contained more than five conversational turns, ranged from 1.13% (Asiana Airlines 214) to 67.3% (PSA 182).

The degree to which flight crew engaged in safety voice changed over time, but surprisingly the degree of safety voice became less overall ($OR = .925$, $Wald(1) = 810.191$, $p < .001$), see Figure 6.2. This was consistent with accidents over time being more frequently attributed to poor crew interaction ($OR = 1.065$, $Wald(1) = 9.387$, $p = .002$). Flight crew were especially

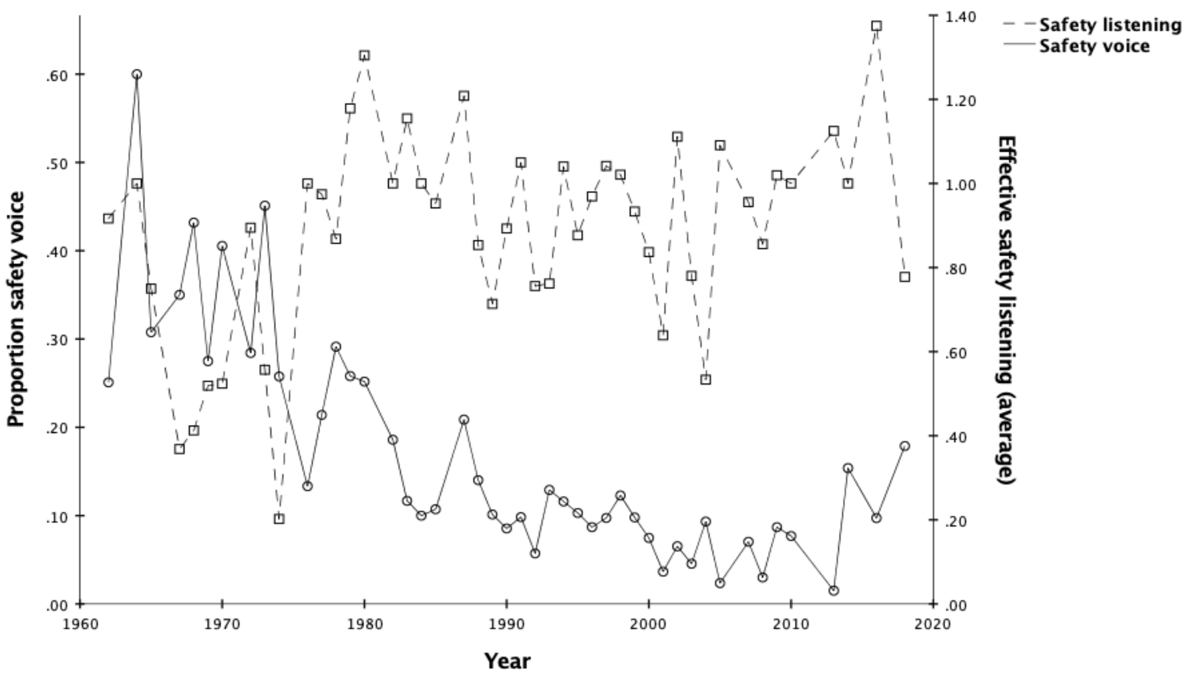


Figure 6.2. Historic trends of the proportion of safety voice and average response to safety voice within CVR transcripts.

less likely to engage in safety voice during historic accidents after the introduction of CRM in approximately 1981[30] (*OR* = .114, *Wald(1)* = 619.810, *p* < .001).

Given the very high number of safety voice acts identified, we describe some of the cases to illustrate that the effectiveness of voice depends upon the constraints of the technical systems, and crew responses to voice. Often, crews voiced concerns after the point at which they could be addressed. For instance, USAF 27, where the co-pilot referred to the potential for bird strike by saying “lot of birds here”. The captain acknowledged (“Lotta birds here”), however the crew were unable to respond quickly enough to the hazard (which air traffic control had neglected to raise awareness of), leading to the loss of 4 crew and 20 passengers. In another case (SAS 751), the first officer voiced several times during an ongoing event (e.g., "We have problems with our engines, please... we need to go back to, ... to go back to Arlanda"), and despite the crew recognising the problem they could not resolve it because the problem

[30] The year 1981 was chosen because CRM programs emerged in the early 1980s (Kanki et al., 2019). Yet, it should be noted that CRM was not simultaneously introduced across airlines.

was an underlying technical issue (ice on the wings). Finally, in another case, Saudia 163, safety voice was repeatedly engaged in (e.g., by the first officer continually raising concerns about smoke in the cabin). The captain and crew responded to this, however a lack of coordination amongst the crew rather than dismissal of safety voice was found to lead to the accident.

Poor safety listening

In general, poor safety listening reduced the overall proportion of safety voice in a transcript ($\beta = -.200$, $F(1,156) = 6.499$, $p = .012$, $R^2 = .040$) and specifically for junior flight crew speaking-up ($\beta = -.212$, $F(1,109) = 5.105$, $p = .026$, $R^2 = .045$). Listening behaviours (n = 1090) tended to be favourable but varied across accidents ($M = .821$; $SE = .022$; $t(1089) = 37.665$, $p < .001$): 82 accidents (e.g., Alaska airlines 261) only saw effective safety listening, 3 only one negative response (i.e., Aviation services, Crossair 498, Martinair 492), and 33 accidents saw repeated poor listening (range: 2-33 times; e.g., Texas International 655). Junior flight crew were listened to less, compared to senior flight crew ($F(1,1088) = 4.590$, $p = .032$, $\eta^2 = .004$). Safety listening became more favourable over time ($F(1,1080) = 26.621$, $p < .001$, $R^2 = .024$), with the introduction of CRM providing a strong historic turning point because listening became more favourable on average after this ($F(1,1080) = 18.142$, $p < .001$, $\eta^2 = .563$), see Figure 6.2.

To illustrate the nature of safety listening, we report on exemplar cases in which voice was ignored or dismissed. For instance, in Kalitta 808 which crashed due to a stall, two voice acts by a flight engineer about low airspeed ("You know, we're not getting our airspeed back there" and "Watch the, keep your airspeed up") were ignored by the crew, who were focussed on identifying the strobe light for landing (e.g., in response to concerns the captain asked "Where's the strobe?"). Similarly, for flight TWA 514 which crashed due to flying at an unsafe altitude, repeated attempts by the first officer to share concerns about the altitude measurement ("I hate the altitude jumping around"; "Gives you a headache after a while, watching this jumping around like that") were not acknowledged by the captain, who was focussed on visually identifying the ground. In other cases, safety voice led to disagreement: during landing in a Metro II aircraft the first office voiced on the landing gear "is it down?", which led to confusion between the captain ("yeah gear's down") and the co-pilot ("No its up") which was not resolved in time. Similarly, in the case of Aeroflot 9981, a co-pilot's request to disengage from a dangerous landing ("No, let's...go around") was dismissed by the pilot ("Why are we going around?"), who then confirmed the action too late ("Tell them "go around").

The role of power distance

Power distance only explained the extent of safety voice, but not safety listening. The proportion of safety voice in a transcript was not predicted by direct effects for the seniority of the voicer ($OR = 1.010$, $Wald(1) = .024$, $p = .877$) and power distance ($OR = .998$, $Wald(1) = .619$, $p = .431$), and as shown in Table 6.3, this emerged due to an interaction-effect between seniority and power distance on safety voice ($OR = 1.015$, $Wald(1) = 10.048$, $p = .002$). Indicating that norms for engaging with seniors shape safety voice, power distance had a negative association with safety voice (Spearman $r = -.034$, $p < .001$)[31], and only predicted safety voice for junior flight crew in low power distance countries ($OR = .990$, $Wald(1) = 7.845$, $p = .005$), but not for senior flight crew ($OR = 1.005$, $Wald(1) = 2.538$, $p = .111$). To illustrate this interaction: junior flight crew were 1.653 times less likely to engage safety voice with a 50-point increase in power distance (i.e., half the scale). Moreover, as illustrated in Figure 6.3, the identified historic decline in the extent of safety voice was especially strong for low power distance countries: a strong interaction-effect existed for power

[31] The relationship between power distance and safety voice was estimated to be quadratic, with the likelihood for a conversational turn containing safety voice = $.393 - .012(\text{PDI}) + 9.859{*}10^{-5}(\text{PDI}^2)$, $F(2,12929) = 35.167$, $p < .001$.

distance and year on the likelihood that a conversational turn involved safety voice (*OR* = 1.003, *Wald(1)* = 98.583, $p < .001$) and the amount of safety voice in a transcript ($F(34,50) = 3.262$, $p < .001$, $\eta^2 = .689$).

Surprisingly, power distance was not associated with poor safety listening to junior flight crew speaking-up ($r = -.041$, $p = .681$), with only a weak association (Spearman's $r = -.071$, $p = .033$) indicating that voice may have been less ignored in high power distance airlines because it involved a more extreme act. Furthermore, safety listening did not explain the effect of power distance on safety voice because no mediation-effect was found in general ($b = .000$, $SE = .002$, *95CI:* -.004 – .005) or for junior flight crew specifically ($b = .008$, $SE = .025$, *95CI:* -.028 – .071), and no interaction-effects existed for power distance with seniority on safety listening ($F(20,866) = 1.297$, $p = .172$, $\eta^2 = .029$) and with safety listening on the proportion of safety voice in a transcript ($F(1,141) = .540$, $p = .464$, $\eta^2 = .004$). However, and consistent with the reduction in safety voice, an interaction-effect indicated that

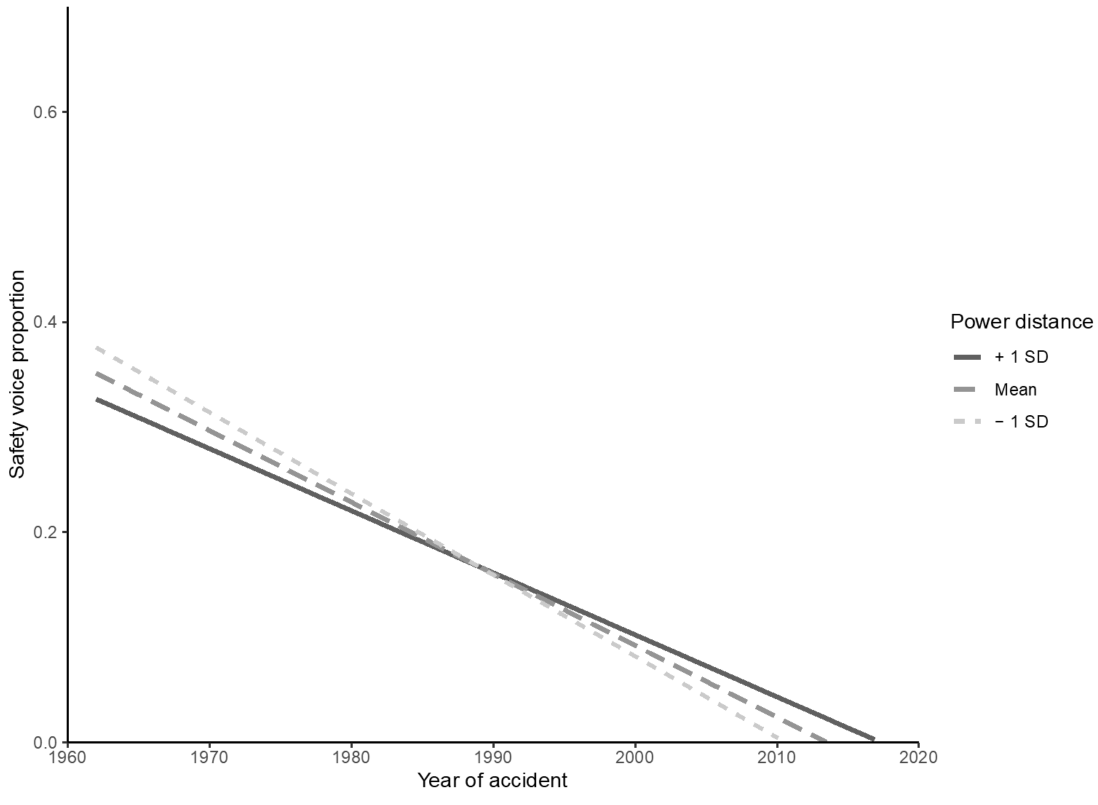


Figure 6.3. The probability of a conversational turn involving safety voice given the year of the accident and airline power distance.

safety listening only became more favourable over time for low power distance airlines (F(22,829) = 2.057, p = .003, η^2 = .052).

Discussion

Through providing the first systematic and behavioural analysis of safety voice prior to aviation accidents, we demonstrated that safety-critical staff nearly always speak-up across hazardous situations. Safety voice was frequently listened to poorly, and this reduced the amount of safety voice in the lead-up to accidents. Confirming previous propositions, power distance explained the extent of safety voice, but no effect was found for safety listening. Moreover, the introduction of CRM training only led to historic changes in safety voice where power distance was low. These findings have important implications for safety voice theory and safety management.

Theoretical implications

We provided the first evidence that people engage in *real* safety voice behaviour during *genuine* accidents, and indicated they do this nearly always across accidents. This is important because the safety voice literature has assumed that accidents can emerge from a lack of safety voice (Kolbe et al., 2012; Noort et al., 2019b; Tucker et al., 2008), yet we indicated that accidents still occurred despite flight crew speaking-up. Thus, in contrast to prevailing thought, we indicate that accidents cannot be assumed to emerge from a lack of safety voice, or that the behaviour is sufficient for avoiding harm. This means that through relying on selective case studies, inquests and studies operationalising hazards (Anicich et al., 2015; Driscoll, 2002; Enomoto & Geisler, 2017; Francis, 2013, 2015; Soeters & Boer, 2000), research has provided insufficient insights on behaviour in the field and wrongly assumed the central problem is an absence of safety voice. Research should therefore be grounded in the analysis of safety voice during actual hazards, and progress from making safety voice more likely to making safety voice more effective (i.e., for preventing harmful outcomes; Bienefeld & Grote, 2012; Kolbe et al., 2012).

Most importantly, we indicated that safety concerns were often ignored or rejected, and this suggests that safety listening may be conceptualised as an essential step in the chain between hazards eliciting concerns, people raising concerns and threats being mitigated (see Figure 6.4). This is important for making safety voice more effective because whilst the safety voice literature has established that anticipated responses from leaders are important (Barzallo Salazar et al., 2014; Bienefeld & Grote, 2014; Edmondson, 1999; Krenz et al., 2020; Nembhard & Edmondson, 2006; Newman et al., 2017; Weiss et al., 2018) it has underdeveloped the role of safety listening. Part of listening effectively to safety voice is responding in constructive ways (e.g., taking action, demonstrating personal interest; Detert & Burris, 2007), which may confirm risk perceptions and enables more voice (Lin & R. E. Johnson, 2015), and we support the generalisation of research on leaders' poor safety listening from controlled environments (e.g., Barzallo Salazar et al., 2014; Weiss et al., 2018) through demonstrating that the degree to which flight crew spoke-up during aviation accidents was lower when concerns were poorly listened to. Thus we enable the application of concepts such as psychological safety (Edmondson & Lei, 2014) and advice taking (Tost et al., 2012) to real accidents, and we suggest future research investigates how safety voice can be made more effective through distinguishing between safety voice and safety listening, and the design of interventions that enable recipients to enact change (Barlow et al., 2019). For instance, through exploring how

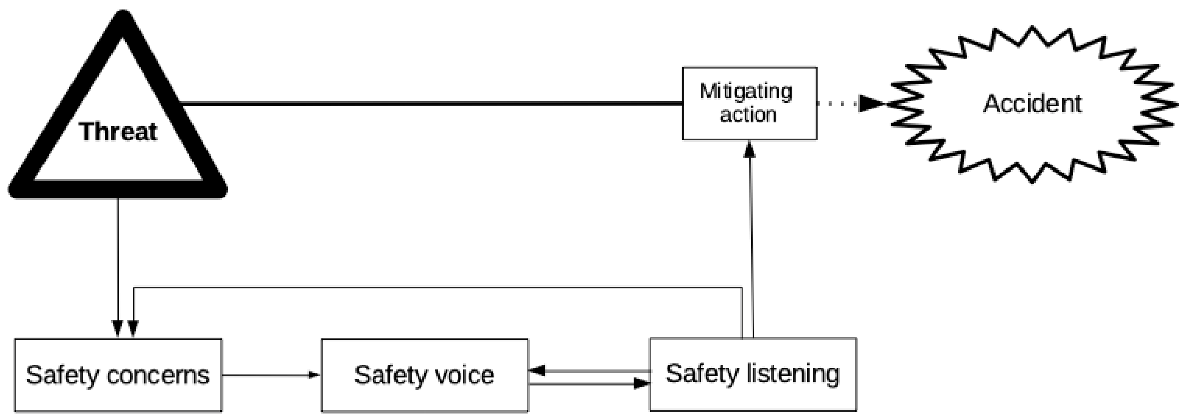


Figure 6.4. Threat Mitigation model of safety voice.
Note: the model highlights that the dysfunctional momentum of threats towards accident (Barton & Sutcliffe, 2009) can be mitigated (dotted line), when threats elicit higher degrees of safety concerns, safety voice and safety listening.

the concept of loss aversion (Tversky & Kahneman, 1981) explains effective listening when people perceive risk.

Our findings are consistent with the notion that the perception of risk provides motivation for sharing situational awareness and initiating decision-making (Barton & Sutcliffe, 2009; Christian et al., 2009; Loewenstein et al., 2001; Slovic, 1987), and support vignette-based and experimental findings indicating that risk is central to safety voice (Noort et al., 2019a; Schwappach & Gehring, 2014c). However, few safety voice studies have assessed risk or delineated leading indicators of accidents (e.g., unsafe acts, or their preconditions; Reason, 1990; Wiegmann & Shappell, 2016). Yet, because we indicated that safety voice is more prevalent during accident than typically established in the literature[32], this indicates a need for outlining how findings from other methodologies (e.g., surveys, interviews, experiments; Noort et al., 2019b) may be mapped unto real hazards in terms of distinct sociotechnical risk factors (Appelbaum, 1997). For instance, future studies may enable the comparison of safety voice and safety listening across hazardous situations through carefully describing how hazard

[32] A one-sample T-test revealed the proportion safety voice was significantly different from literatures indicating people only speak-up in 44% of the cases in which they are concerned ($t(170) = 66.494$, $p < .001$).

characteristics (i.e., in terms of technical or physical properties and levels of risk) elicit visceral states which are difficult to recall or forecast (Loewenstein et al., 2001; Noort et al., 2019a).

In addition, we showed that whilst safety voice occurred across accidents, the amount of safety voice varied across transcripts. This is important for safety voice theory because it confirms that factors beyond physical risk influence the degree to which people speak-up about safety (Noort et al., 2019b) and thus, whilst it is essential to increase the effectiveness of the behaviour, scope remains for increasing the degree to which people speak-up. In particular, whilst leader behaviours (e.g., power distance, leadership styles) have been proposed to cause accidents through reducing safety voice (Gladwell, 2008; Kwon et al., 2016; Liang et al., 2012; Morrison, 2011, 2014; Wilkinson et al., 2020), we provided the first direct and systematic evidence that social structures can reduce safety voice during actual accidents. Additionally, we provide an important nuance to the power distance proposition for accident causation through highlighting that power distance reduces safety voice, but not through leaders listening more poorly in high distance environments. Thus, we confirm research indicating that power distance contributes to accident rates (Enomoto & Geisler, 2017; Soeters & Boer, 2000), evidence the generalisability of findings on the individual level construct of power distance orientation (Botero & Van Dyne, 2009) to influences on safety voice within safety-critical teams, and indicate the need to investigate how power distance and safety listening independently reduce safety voice. Moreover, we make tractable the investigation of established safety voice antecedents (e.g., leaders using inclusive language; Weiss et al., 2018) and interventions (e.g., education-based training; O'Donovan & McAuliffe, 2020) during real-life hazards. Finally, this contributes to the wider safety management literature (e.g., risk perception, safety citizenship, safety culture; Didla et al., 2009; Guldenmund, 2000; Slovic, 1987) through indicating that the investigation of sedentary risks (e.g., on behaviours that maintain relatively stable risk) may be supplemented by the investigation of safety voice and safety listening because it provides access to social mechanisms explaining how people communicate during emergencies.

Fourth, we indicated that the introduction of CRM provided a good explanation of historic trends in safety voice and safety listening, yet we found that safety voice declined over time. This is surprising because it contradicts the literature that suggests CRM improves speaking-up (Kanki et al., 2019), but it may be explained by CRM improving safety listening (e.g., through increases psychological safety; Edmondson, 1999) and thus reducing the need for repeated safety voice (i.e., because cooperative relationships increase shared situational awareness; Driscoll, 2002; Foushee, 1984) or even preventing accidents (and thus the inclusion in the dataset). This would support the use of CRM training, and through providing the first evidence on reduced effectiveness of CRM in higher power distance contexts, we indicated a need for research to improve CRM training across cultural contexts.

Finally, investigating safety voice and safety listening through a cultural lens can extend safety voice theory through the identification of additional cultural predictors of safety voice. Safety voice research has rarely done this (Noort et al., 2019b), but this would be valuable for the design of new interventions. Future research may identify cross-cultural differences in safety voice due to face-saving (Mao, 1994), global differences in leadership values and practices (House et al., 2004), or other national culture dimensions (e.g., individualism, uncertainty avoidance, masculinity, long-term orientation; Hofstede et al., 2010). Yet, it may prove more optimal to develop the concept of safety voice as an integral activity to organisational politics (Antonsen, 2009) and sense-making on risk (Douglas, 1992; Weick, 1995). These approaches describe how cultural processes emerge in response to challenges for dealing with risk, and adopting them may extend existing perspectives (e.g., highlighting that safety voice results from voice climate; Morrison et al., 2011) through indicating how safety voice and safety listening dynamically constitute safety culture. For instance, through longitudinal investigations on the sense-making process through which the behaviours lead to institutional change.

Practical implications

Our results have practical implication for safety management and safety-critical teams. First, unlike previously assumed (Kolbe et al., 2012; Noort et al., 2019b; Tucker et al., 2008), safety voice occurs during accidents but its effectiveness for avoiding harm needs to improve: we indicated a gap between safety voice and the mitigation of harm. This means that whilst safety voice is necessary for avoiding accidents, it provides incomplete protection (e.g., in terms of Reason's Swiss Cheese model; Reason, 1990) without practitioners recognising and responding appropriately to concerns raised (e.g., through engaging in open conversation, taking action). Thus, whilst safety voice contributes to the mitigation of risk, steps need to be evaluated for increasing the effectiveness of safety voice, for instance through improving safety listening, and this should be incorporated into training programs such as CRM (Kanki et al., 2019).

Second, our findings support the scope and benefit of CRM training programs. This is because variation in the degree of safety voice during accidents indicates interventions may improve the behaviour, and the historic introduction of CRM led to better safety listening and, as argued above, safety voice. However, whilst research has indicated the impact of cultural norms on safety behaviours and accidents (Merritt & Helmreich, 1996; Reader et al., 2015; Soeters & Boer, 2000), we indicated that CRM training remains insufficiently tailored to high power distance environments. This is especially pressing for safety management in these environments because research indicates that accidents are more likely where norms do not support egalitarian interactions (Enomoto & Geisler, 2017). After research increasing CRM's overall effectiveness (Helmreich et al., 1990), the next phase of CRM implementation should therefore tailor training programs to specific environments.

Limitations

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

First, the quality of the dataset is dependent on included CVR transcripts, the condition of the source files after accidents occurred, and the standard of transcription (Sassen, 2005). Included transcripts were available at the online databases and written in English, and other transcripts may have been missed. However, the dataset incorporated approximately 15% of commercial and corporate aviation fatalities since 1962 and thus the data provides substantial coverage of known cases. Not all original audio files were accessible, and we needed to assume reasonable transcription accuracy. We suggest this is appropriate because providing accurate transcripts is in the interest of accident investigations, and transcription uncertainties were indicated in the transcripts (e.g., 'unintelligible'). Future research may enhance the dataset through extending the number of transcripts (e.g., new accidents, or from alternative sources), or directly testing the transcription quality.

Second, the analyses only enable tentative conclusions on the occurrence of safety silence and outcomes. We demonstrated a high degree of safety voice across accidents and because people speak-up in response to perceived hazards (Noort et al., 2019a) it is highly probable that flight crew spoke-up because they perceived risk during the accidents. However, whilst conclusions on the extent of safety voice were possible, the absence of safety voice does not readily constitute safety silence (i.e., flight crew may not speak-up because they are not concerned; Noort et al., 2019a). Future research may investigate text-based measures for assessing safety concerns in flight crew speech and apply these to establish conclusions on safety silence. Additionally, normal flights and near-misses were not included in the dataset and this means that conclusions are not straightforward on the extent that safety voice would have avoided harm. Because of this the attribution of blame is not only undesirable, but invalid.

Safety voice theory may advance considerably through establishing how safety voice enables the avoidance of harm, and this may be optimally achieved through triangulating the CVR dataset with near-misses and safety performance data (Blanco et al., 2009).

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

Finally, the appropriateness of using Hofstede's dimensions has been debated (Hofstede, 2002; McSweeney, 2002). People within countries display a broad range of psychological tendencies (Kitayama et al., 2009), and whilst cultures remain relatively stable, 172 accidents may not reflect the heterogeneity of cultures. We suggested that national-level data may be used as a proxy for power distance on the flight deck, and through presenting variation in the degree to which people raise concerns across 14,128 conversational turns from pilots from 42 countries we provide a first step in this direction. The literature may further reduce potential biases from homogenous samples through replicating these findings for other hazards and industries.

Conclusion

Safety voice is theorised as an important mitigating factor for maintaining safety, but behavioural research during actual hazards has been scant. We showed that historic accidents that posed fatal risk elicited high levels of safety voice, but variation in the degree to which safety voice dominated conversations for junior flight crew was reduced by poor safety listening and high power distance. We showed that the literature cannot assume that safety voice does not occur during accidents or that it is sufficient for avoiding harm: accidents occurred despite high degrees of safety voice and a need exists to improve the effectiveness of safety voice through enabling people to listen more effectively to safety voice. This is especially important for tailoring CRM training programs to high power distance environments. Across sociocultural contexts, people mitigate hazards through engaging in conversation with others, and the field needs to incorporate how people enact safety voice because raising and listening to safety concerns provide unique challenges for avoiding accidents.

CHAPTER 7:
DISCUSSION AND OUTLOOK

Safety voice (Barton & Sutcliffe, 2009; Conchie et al., 2012; Tucker et al., 2008) contributes to the prevention of injuries and accidents through people making sense of how perceived risk from hazards may be mitigated. The safety voice literature has stressed that interventions are needed because coordination on safety is challenging and well-intended individuals often fail to speak-up upon encountering safety threats (Conchie et al., 2012), with safety silence being identified as contributing to tragic accidents across safety-critical industries (e.g., healthcare, energey, aerospace; Bromiley & Mitchell, 2009; e.g., Moorhead et al., 1991; NTSB, 1978; Reader & O'Connor, 2014). Thus, and by drawing predominantly from safety management (e.g., safety culture; Guldenmund, 2000) and voice research (e.g., employee voice; Hirschman, 1970; Morrison, 2011) and adopting positivist approaches (Kenny et al., 2020), the literature has treated safety voice as a key factor in understanding accident causation (e.g., by using illustrative examples of tragic mishaps; Bienefeld & Grote, 2012) and has indicated that the likelihood of people raising safety concerns can be increased by favourable antecedents (e.g., psychological safety, favourable leader behaviours; Manapragada & Bruk-Lee, 2016; Weiss et al., 2018). In this thesis, I contributed towards safety management by providing a unified and distinct concept of safety voice, and highlighting that safety voice may be investigated within sociotechnical approaches (Appelbaum, 1997; Leveson, 2002; Reason, 2000) because I provided the first systematic evidence that the phenomenon is highly ecological and situated in the social and technical characteristics of hazardous scenarios.

To date, the safety voice literature has remained disintegrated and has not clarified to what extent safety voice is relevant for understanding why accidents occur. In particular, whilst safety management theories have incorporated concepts and measures that capture safety voice (e.g., Julian Barling et al., 2002; Didla et al., 2009; Guldenmund, 2000; Neal & M. A. Griffin, 2004; Zohar, 1980), it remains unclear how safety voice should be optimally conceptualised, assessed and intervened in. Accordingly, in this thesis, across four articles, I undertook fourteen studies ($n_{participants} = 1{,}222$, $n_{archival} = 220$) that contributed insights towards three research questions (see Table 1.1): *1) What is the behavioural nature of safety voice?; 2) What is the optimal way to investigate safety voice behaviour?;* and, *3) To what extent do interventions for promoting safety voice and reducing safety silence need to be tailored to the behavioural nature of safety voice?*

In summary, these studies evidenced that safety voice is highly situated in hazards and social interactions: the behavioural nature of safety voice can be conceptualised in ecological terms across levels of analysis (e.g., hazards, teams, organisations, Chapter 3), and as i) contingent upon hazardous situations in terms of temporal development and risk (e.g., risk eliciting safety concerns; Chapters 3-6), ii) manifested in the degree to which concerned individuals engage in safety voice speech that varies according to contextually relevant themes (i.e., safety knowledge and motivations) and the temporal progression of hazardous scenarios (Chapter 5), and iii) dependent on favourable behaviours from others (e.g., safety listening) for mitigating accidents in hypothetical and real hazardous scenarios (Chapters 4-6). As I highlighted within the four presented articles, these findings contribute in important ways to research on safety voice and safety management and have contributed towards the development of the Threat Mitigation Model of safety concerns, safety voice and safety listening (see Figure 7.1). Below, I briefly summarise specific contributions to the safety voice and safety management literatures before zooming out to the broader implications for theory, method and interventions.

Overview of specific contributions

First, the proposed ecological framework for safety voice provides the first systematic delineation of safety voice across levels of analysis. This is important because the systematic literature review revealed that safety voice research is grounded in distinct safety (Aydon et al., 2016; Guldenmund, 2000; Haavik et al., 2015; H. L. Johnson & Kimsey, 2012; Lyndon, 2008; Reason, 1990) and communication research (e.g., Hirsschman's EVL framework, Morrison's employee voice; Barton & Sutcliffe, 2009; Hirschman, 1970; Morrison, 2011; Okuyama et al., 2014; Tucker et al., 2008). These have previously conceptualised multi-level models (e.g., the Swiss Cheese Model; Reason, 1990), yet safety voice research has not clearly delineated findings across levels of analysis (Morrow et al., 2016; Okuyama et al., 2014). Thus, by providing an integrated concept that is delineated across levels of analysis, I enable the application of Erez and Gati's conceptualisation of multiple layers of analysis (Erez & Gati, 2004) to the safety voice literature, and the investigation of safety voice within systems perspectives of safety management (Leveson, 2011; Reason, 2000; Wiegmann & Shappell, 2016) that incorporate immediate (e.g., hazard and individual antecedents) and leading indicators of safety voice (e.g., antecedents related to the institution and external context).

Furthermore, previous research had not evidenced how safety voice behaviour is related to the perception of risk (i.e., to my awareness only one study has established how risk perceptions shape safety voice; Schwappach & Gehring, 2014c) or the progression of time (i.e., studies have not delineated phases of hazards, or associated these with interventions; Farh & Chen, 2018; Krenz et al., 2020), and had not provided a systematic analysis of safety voice behaviour during actual accidents (Krenz et al., 2020). Thus, by evidencing that safety voice is highly contingent upon the characteristics of hazards (e.g., risk, time, others' behaviour) in experimental scenarios and the field (i.e., aviation accidents), I underscore the importance of the literature to ground safety voice research in real accidents and assess risk perceptions to enable insights into the generalisability to high-stakes hazards. In particular, the literature had provided valuable but limited evidence on the manifestation of safety voice in speech (i.e., studies have provided limited insights into the words used; Krenz et al., 2019), and no systematic evaluation existed on safety silence and safety listening. I demonstrated that conceptualising safety silence based on the extent to which people that perceive risk engage in distinct types of safety voice speech is essential for designing successful interventions. Moreover, conceptualising safety listening is essential for understanding how safety voice can be effective in mitigating accidents.

Furthermore, in terms of methods, I identified that the literature has produced valuable insights through post-hoc and self-report measures (see Chapter 3; e.g., the themes people use to describe their safety voice/silence; Morrow et al., 2016) and would benefit from direct behavioural observations (e.g., through experiments or analysis of archives containing field-based safety voice) that make tractable the testing of new hypotheses (see Chapters 4-6). In particular, few paradigms existed for conducting laboratory experiments on safety voice (Barzallo Salazar et al., 2014; Hodges, 2018) and their limitations were unclear. Thus, by evaluating the Walking the Plank paradigm I provided a new experimental paradigm and demonstrated that safety voice experiments can overcome the literature's methodological and ethical challenges for assessing the behavioural nature of safety voice, overcoming the reliance on memory and imagination and establishing causal relationships (see Chapters 4 and 5). In addition, I illustrated how the investigation of safety voice may obtain empirical and ethical insights into the nature of the behaviour during real-life hazards by investigating historic accidents (see Chapters 2 and 6), and underscored the importance by revealing that real hazards elicit substantially more safety voice compared to the average levels of safety voice found in the literature (i.e., ±44%; see Chapters 3 and 6).

Finally, O'Donnovan and McAuliffe (2020) indicated that the safety voice literature has only had mixed success for safety voice interventions, and I revealed that this may be explained because the prevailing literature has rarely accounted for the extent to which interventions are shaped by the relationship between safety voice and hazardous situations (e.g., risk perceptions, actual hazards, temporal progression of hazards). I revealed that this is essential for improving intervention success because safety voice can be conceptualised as ecological and contingent upon hazardous situations: safety voice measures and interventions need to account for the degree to which people are concerned about hazards (i.e., to establish safety silence) and how others listen to safety concerns, and tailor designs to the intended context and stage of hazards (e.g., national culture, conceptualisation stage; see Chapters 4-6).

These specific contributions to the literature provide an integrated concept for safety voice and a new direction for research (e.g., closing gaps revealed in the ecological framework, directly observing behaviour to characterise the behavioural nature of safety voice, integrating related concepts on voice and harm-prevention, evaluating the extent to which research has accurately operationalised safety silence, see Chapter 3-6). Moreover, they indicate broader contributions to theory, method and the application of safety voice to safety management.

Implications for theory and methods

In addition to the specific contributions identified and summarised in the articles, an analysis of these contributions highlights more broad ranging implications for theory and method related to safety voice and safety management. Thus, and to avoid repeating the previous chapters, here I zoom out and evaluate the wider implications of my findings for conceptualising and assessing safety voice. Based on this, I suggest new directions for research.

Safety voice provides a unique concept, but draws on intersecting ideas

By systematically reviewing research capturing safety voice behaviour, I demonstrated that the concept of safety voice draws upon research across diverse domains (i.e., safety, communication and team-working) and methodologies (e.g., ethnography, surveys, simulations). The specific focus on safety voice behaviour enabled the synthesis of evidence, the provision of an integrated definition, a proposal for a unified framework and a conceptual comparison with employee voice. This is important for advancing safety voice theory because previous systematic literature reviews did not outline the range of theoretical backgrounds (Morrow et al., 2016; Okuyama et al., 2014). This indicates that distinct voice and safety domains draw on intersecting ideas (especially on the act of speaking-up) and I therefore support the argument by Wilkinson and colleagues (2020) that a conversion of terminology and models should be evaluated. However, and confirming Tucker and colleagues' (2008) proposition, I showed that integration of concepts may not be straightforward because safety voice provides unique scope (see Chapter 3) and data (i.e., Chapter 4 highlighted that employee voice was not associated with safety voice) compared to employee voice (Morrison, 2011, 2014). For safety voice theory, this means that whilst frameworks may be proposed by drawing on related research, I indicate a conceptual gap: it is unclear how concepts drawing on similar ideas can provide different empirical findings.

I contributed towards conceptualising the distinction between employee voice and safety voice, and this may feed into a Delphi study (Linstone & Turoff, 1975) to build an integrated voice model explaining variation across voice concepts. For instance, and arguably, future research may contribute a single model that clarifies how distinct voice concepts emerge from variation in terms of the context (e.g., hazards, organisations), content of speech (e.g., expressed themes; promotive or prohibitive voice, Liang et al., 2012), motive (e.g., Brinsfield, 2013; Manapragada & Bruk-Lee, 2016), voicer (e.g., bystanders; P. Fischer et al., 2011), recipients (e.g., authority figures or equals; Kaposi, 2017; Wu Liu et al., 2010), timing (e.g., hazard stages, duration until voice occurs; Krenz et al., 2020) and outcomes (e.g., Bashshur & Oc, 2014).

Thus, and addressing the need identified by Wilkinson and colleagues (2020) for evaluating and integrating intersecting ideas on voice, Figure 7.1 provides a model for how voice can mitigate threats (e.g., poor working conditions, unsafe operation) in order to avoid failures (e.g., accidents, organisational decline, etc.). This conceptual model integrates concepts and

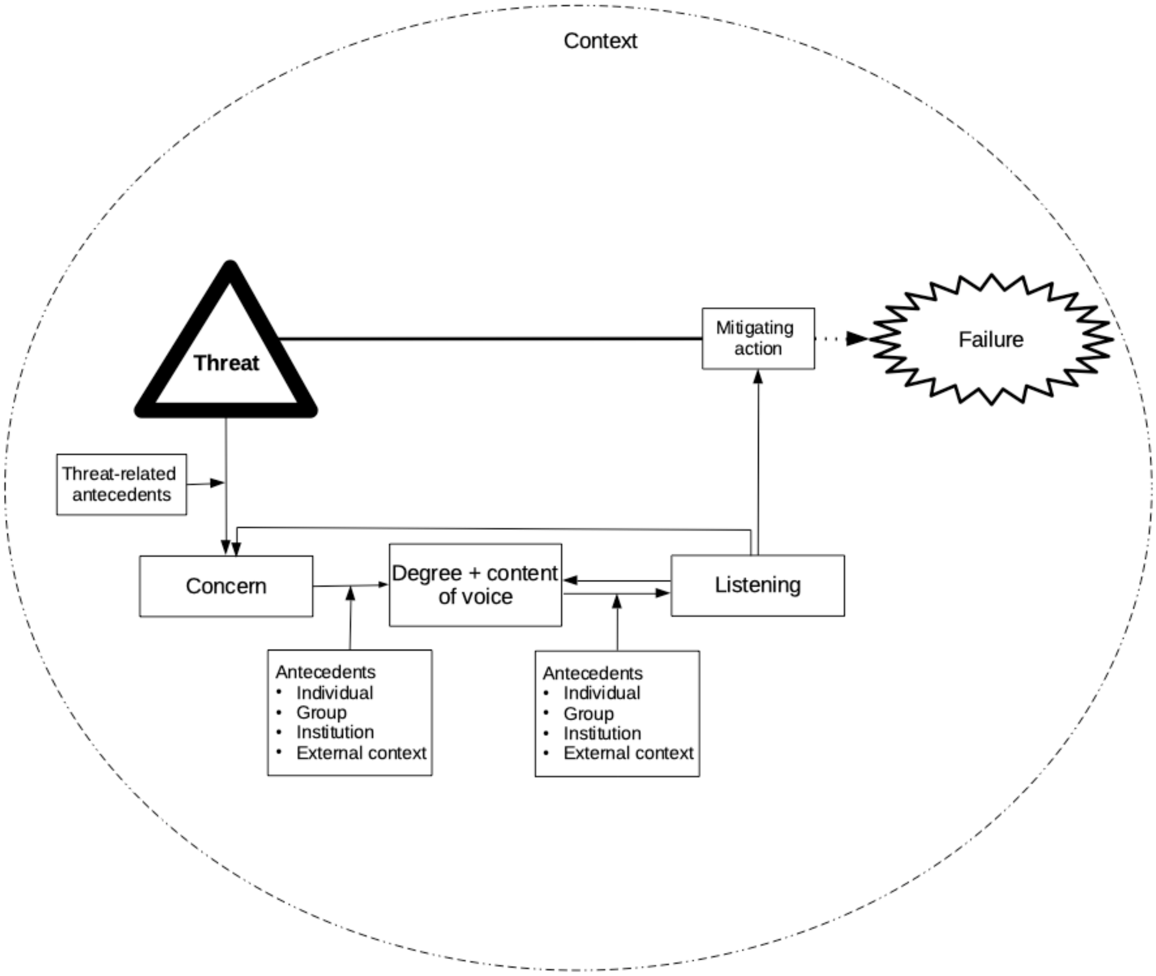


Figure 7.1. Threat Mitigation Model of concerns, voice and listening.
Note: this model extends the models developed in Chapters 1-6 to voice failures more broadly. It highlights that, within a given context (e.g., hazardous scenarios, organisations), the momentum of threat towards failure (bold line; Barton & Sutcliff, 2009) can be mitigated through effective commination that may itself be mitigated when antecedents reduce concerns, voice and listening. For the purpose of a clear presentation, variables are summarised.

evidence developed and outlined in the previous chapters, highlights the advancements in knowledge compared to the safety voice literature, and underscores the essential role of safety concerns (see Chapter 4-5), safety-related content (see Chapter 5) and listening by others (see Chapter 6). Thus, it adds to safety management by highlighting how principles of Reason's Swiss Cheese model (e.g., that safety management provides different layers of threat mitigation; Reason, 1990) can be applied to voice behaviours, with concerns, voice and listening providing distinct, and stacked, factors for understanding how failure (e.g., accidents) may be mitigated. Thus, this model makes tractable new hypotheses on i) the extent to which intersecting ideas in the broader voice literature may be integrated (e.g., motivations for silence; Brinsfield, 2013; Manapragada & Bruk-Lee, 2016), ii) the effects of safety listening on mitigating threats (Barlow et al., 2019; Burris et al., 2013; A. Jones & Kelly, 2014), and iii) the distinct impacts of concerns, voice and listening on accident causation.

The behavioural nature of safety voice is situated

Revealing the ecological nature of safety voice contributes towards understanding safety voice as a situated phenomenon. I revealed that safety voice varies according to at least 32 second-order antecedents across "Erez and Gati's (2004) levels of analysis for the individual, group, institutional and external context" (Noort et al., 2019b, p.383). This underscores that the behavioural nature of safety voice is contingent upon individual and contextual variables (Okuyama et al., 2014). By delineating previous findings, I showed that this situated nature of safety voice can be conceptualised in ecological terms (i.e., as related to variables across levels of analysis). For instance, and confirming the conclusion in the doctoral thesis of James Detert (2003), Chapter 3 indicated that leadership styles provide individual (e.g., perceived costs of voicing concerns; Maxfield et al., 2013), group (A. Jones et al., 2016) and institutional level antecedents to effective safety voice (e.g., hierarchy; Noland & Carmack, 2015), and I added to this by revealing that leadership shapes safety voice (i.e., safety listening, power distance) during real hazardous situations (i.e., during aviation accidents; see Chapter 6).

This is important for safety voice theory because it means that safety voice may be conceptualised in holistic (i.e., emphasising that behaviour is situated in context and history) rather than atomistic terms (i.e., reducing behaviour to self-contained components; Fay, 1996) and be shaped through Erez and Gati's (2004) dynamic feedback processes (e.g., between antecedents across levels of analysis, or safety voice, concerns and listening). Yet, whilst multilevel models exist for safety management (e.g., Reason, 1990; Shappell & Wiegmann, 2000) and dynamic feedback processes are shown to impact negatively on employee voice (Lin

& R. E. Johnson, 2014), few safety voice studies have clarified how their variables (e.g., hazards, voice, antecedents) are delineated across levels of analysis (cf. Nembhard, Yuan, et al., 2015) and time (cf. Krenz et al., 2020). This is important because without a situated understanding of safety voice, mixed success for interventions may not be understood (O'Donovan & McAuliffe, 2020), variation in the degree or content of safety voice and silence may not be explained (e.g., because contexts provide cues on risk; see Chapters 5 and 6), and accident investigations would be ineffective. For instance, in terms of recognising thematic variations in speech (see Chapter 5) or describing how voice and silence emerged during a hazardous scenario (e.g., from fatal risk posed by aviation accidents, poor safety listening; see Figure 7.1).

Thus, by indicating the ecological nature of safety voice, I outline the scope for investigating safety voice within complex (i.e., dynamic and multi-level) feedback processes that can impact on safety (e.g., when the initial dismissal of concerns reduces subsequent safety voice and sets precedent) and ultimately lead to accidents such as with United Airlines Flight 173 (NTSB, 1978) and Deepwater Horizon (Reader & O'Connor, 2014). For instance, research on safety citizenship (Reader et al., 2016) indicates that safety-related behaviours may be explained by variables across levels of analysis (e.g., citizenship behaviour, perceived organisational support), and research may explain how accidents are enabled by identifying causal priority (Boyce et al., 2015) between voicing and listening. This is consistent with systems and sociotechnical approaches to safety (Appelbaum, 1997; Leveson, 2002; Reason, 2000) and structural theories of culture (e.g., Berger & Luckmann, 1966; Schein & Scheiner, 2016) that together describe how, over time, feedback on effectiveness enables behaviour to become learned, legitimised and normative for subsequent behaviour (i.e., become normalised; Vaughan, 1999). Thus, by providing an ecological conceptualisation, and evidencing the unfavourable effect of poor safety listening, I enable the interpretation of accidents such as with the Challenger space shuttle (Moorhead et al., 1991; Vaughan, 1996) in terms of concerned individuals (e.g., Thiokol engineers) learning to be silent based on previous safety voice: poor safety listening indicated that safety voice was considered inappropriate and ineffective for maintaining good relationships (also see: Edmondson, 1999).

Furthermore, the situated nature of safety voice may be optimally understood as reflecting the fact that safety voice emerges within the interaction between individuals, others and hazards (also see Figure 7.1). Previous research has provided ample evidence that safety voice is shaped by the interaction between individuals and others (A. Jones et al., 2016). However, the literature has provided surprisingly little behavioural evidence on the mechanisms through which hazards shape safety voice (with self-reported data indicating people tend to explain safety voice due to the higher impact of hazards; Schwappach & Gehring, 2014c) and has given little consideration to whether people understand the threats they encounter (e.g., Barzallo Salazar et al., 2014), and has not assessed safety voice during actual risks (see Chapter 6; Krenz et al., 2020), or how others listen to safety concerns during actual hazards. Thus, it appears that the literature has assumed that hazards elicit safety voice, whilst studies have not assessed safety concerns (cf. Schwappach & Gehring, 2014c).

Yet, I indicated that this assumption, though intuitive[33], is only partially warranted. I evidenced that people perceived risk based on available information in their environment (e.g., the maximum load of the plank in the Walking the plank paradigm) and raised more safety concerns when safety margins became smaller (Chapter 4), when harm was more proximal in time (Chapter 5), or when the actual probabilities of harmful outcomes were very high (Chapter 6). This suggests that the extent to which people engage in safety voice depends on

[33] A repeated question in response to the presentation of the early findings of this thesis was whether people might simply not raise safety concerns because the risks are too minor. This explanation has intuitive appeal: why, after all, would someone raise concerns about *safe* situations? However, as discussed here, the data indicates that a more nuanced understanding of the role of risk is appropriate.

technical hazard characteristics. Nevertheless, I provide an important nuance to this assumption that hazards elicit safety voice by showing that actual risk is neither necessary, nor the sole explanation for safety silence. That is, the perception of risk was enough for the Walking the Plank studies that presented no actual risk, and the extent to which people engaged in safety voice during historic aviation accidents remained high but varied due to power distance and safety listening. Together, this contributes two important theoretical insights.

First, it indicates that safety voice is only loosely contingent upon actual risk because people vary in their risk perceptions (e.g., due to a lack of awareness, or different attitude towards the hazard; Slovic, 1987). Second, whilst to the best of my awareness research has not done this to date, it implies that safety voice may be conceptualised as situated sensemaking (e.g., Weick, 1995, 2010) that emerges from the interaction between a person (i.e., safety concerns, safety voice), other people (i.e., safety listening) *and* hazardous situations (i.e., in terms of the sociotechnical and systemic factors providing risk; Appelbaum, 1997; Reason, 2000). This triangular relationship (i.e., self-other-hazard) is consistent with the ecological nature of safety voice (i.e., it emphasises dynamic interaction over time) and theories on the social construction of knowledge and risk (Bauer & Gaskell, 1999; Berger & Luckmann, 1966; Searle, 1995; Turner & Gray, 2009).

Moreover, it provides a novel perspective on safety voice that moves beyond the literature's tendency to design research based on positivist and individualistic philosophies that reduce the importance of historic and situated factors (Kenny et al., 2020; Turner & Gray, 2009). Such a perspective on safety voice may explain how accidents emerge from the dynamic formation, justification, contestation, and prioritisation of opposing risk perceptions (e.g., safety risks versus interpersonal risks; Edmondson, 1999; Slovic, 1986) and enable new hypotheses on the dynamic between safety concerns, safety voice and safety listening. More broadly, for the field of social psychology, this may provide an alternative reading of Milgram's seminal findings on obedience (Blass, 1999; Milgram, 1974) whereby participants' risk perception is altered through the experimenter providing repeated instructions: this may have reduced the rational likelihood of the participant being correct (Krueger & Massey, 2009) and delegitimised participants' safety concerns through poor safety listening.

Safety voice and safety silence both involve nuanced speech

By evaluating the model for the manifestation of (un)concerned voice and silence (see Chapters 4 and 5) I showed for the first time that safety silence can be scaled based on the extent of safety voice speech for concerned people. This highlighted that the words used reflect the hazardous situation in terms of the stages of hazards (e.g., when harm is a concept versus imminent), available information (e.g., maximum load of the plank) and individuals' motivations for preventing harm, and it advances safety voice theory by indicating that the behavioural manifestation of safety voice and safety silence i) can be investigated within a single continuous conceptualisation for distinct themes in speech, and ii) reflects the process of situated sensemaking discussed above (i.e., capturing the dynamic of safety concerns, safety voice and safety listening as leading indicators of accidents). This is important because, and despite previous research having defined safety voice as communication (Tucker et al., 2008), very few studies have investigated safety voice in terms of the content communicated through speech (a notable expection is provided by Krenz et al., 2019). This is surprising, and my findings underscore the importance of investigating safety voice in-situ because the content for speech is provided by cultural and historic contexts (e.g., Gillespie & Cornish, 2010).

Future research should move beyond investigating the occurrence of safety voice and address how trends in its content across situations enable a more optimal mitigation of accidents. This emphasis on the content of voice corresponds to Mumfords' (2015) and Kurzon's (2007, 2011) conceptualisations of the extent to which themes appear in speech, and it would advance the safety voice literature because it could i) reduce the literature's tendency to investigate perceptions of behaviour (i.e., ±76% of safety voice studies, see Chapters 3-4) by focussing on what people actually say and do, and ii) avoid dichotomous operationalisations (Manapragada & Bruk-Lee, 2016; Tucker & Turner, 2011) by making tractable hypotheses on the extent to which accidents can be prevented by the range of ways that hazards manifest in the extent that people raise distinct safety voice content (e.g., as distinct themes in speech, on continuous scales; see Chapter 5). In addition, research should investigate speech for safety listening as words used by leaders impact on safety voice (Weiss et al., 2018).

Research on natural language processing indicates that semantic computations can be performed on the meaning of words and sentences (e.g., “king” – “male” = “queen”; Pennington et al., 2014; Vylomova et al., 2016) and this may be applied to translate safety voice dictionaries meaningfully across hazardous situations (e.g., informative voice – “weak plank” + “uncleared runway”) to understand how safety voice content, or the lack thereof, contributes to accidents. For instance, in this way the proposed safety voice dictionaries may be ‘translated’ to accident investigations in healthcare and energy (e.g., if speech data is recorded during accidents) and the investigation of safety concerns in patient complaint letters (Gillespie & Reader, 2016; Reader et al., 2014). It may be particularly interesting to explore whether a system may be developed to capture and analyse language automatically in order to warn safety operators in real-time about withheld concerns during medical procedures (e.g., surgical operations), on the flight deck or in online environments such as social media or internet fora (e.g., intelligence operations aiming to capture threats early).

Safety silence may emerge when social factors trump technical factors

Indicating the importance of risk perceptions for safety voice (Loewenstein et al., 2001; Slovic, 1987), I showed that people engaged in safety voice in response to risk (as perceived and evident post-hoc from outcomes), but withheld safety concerns when social demands were unfavourable (e.g., responsibilities, norms and values for engaging with others, poor safety listening). This confirms the value of previous research on safety voice antecedents (Barzallo Salazar et al., 2014; Burris, 2012; Duan et al., 2017; Tucker et al., 2008) for establishing safety voice in the laboratory environment, and confirms previous research suggesting that social motivations play a vital role in voice behaviours (e.g., Edmondson, 1999). Moreover, I extend this research because I showed that this effect remained when participants were more concerned (Chapter 4) or were poorly listened to during high-risk accidents (Chapter 6). This supports the notion that safety silence emerges from a poor expected utility from speaking-up (Murphy & Dingwall, 2007b) and indicates that the act of safety silence may be rational (Krueger & Massey, 2009) and prosocial.

Classic social psychological research indicates that undesirable behaviours can emerge because most people are prosocial (van den Bos & Lind, 2013; van den Bos, van Lange, et al., 2011): most people intend to engage in good behaviour and process social information on how they are expected to act on perceived risk (e.g., intervene, conform to social norms, obey authority figures; Asch, 1956; Darley & Latane, 1968; Milgram, 1974). I did not present evidence indicating that participants in this thesis held prosocial values (for an illustation on how social values may be evaluated in relationship to undesirable behaviour, see: van den Bos, van Lange, et al., 2011), yet my findings are consistent with participants processing social information in order to decide on appropriate and socially desirable action (e.g., less safety voice for poor safety listening and high power distance).

To the extent that this proposition holds (i.e., voice and silence have also been conceptualised as emerging from pro-self motives; Brinsfield, 2013), arguably, safety silence emerges because acting in prosocial ways can be more important to individuals than preventing harm, even to the point where harms are personal (e.g., air crashes). For instance, because they downplay the risk of dying but consider voice socially inappropriate (e.g., due to power distance norms, see Chapter 6; Hofstede et al., 2010). This suggests that accident analyses and investigations should investigate the social motives that people had for engaging in safety silence. Furthermore, this proposition is significant because it indicates that safety voice research may contribute insights to classic social psychological paradigms (Asch, 1956; Latane & Darley, 1968; Milgram, 1974) by indicating the extent to which people are willing to sacrifice their own safety when the situation demands it (i.e., Asch's conformity studies involved no threats to physical safety, Milgram's obedience and Latane and Darley's bystander effect studies involved safety threats that were apparent and victimised others). Whilst social psychological research has informed safety management theories that incorporate safety voice behaviours (Didla et al., 2009; Guldenmund, 2000; Neal et al., 2000), I am aware of little explicit conceptual exchange between the safety voice literature and classic social psychological research (cf. Bienefeld & Grote, 2012; Pian-Smith et al., 2009; Weiss et al., 2018). Future research may adopt a version of the Walking the plank paradigm whereby participants walk across the plank alongside the research assistant to assess and conceptualise the extent to which safety silence can involve self-sacrifice.

A need exists for a broader methodological toolbox

Finally, the findings underscore the importance of expanding the methodological toolbox for investigating safety voice, and safety-related behaviours more broadly. I revealed a need to address existing shortfalls for assessing safety voice because most studies (i.e., approximately 76%, see Chapter 3) rely on data that emerges from methods that ask people to report on previous safety voice behaviour, or to imagine hypothetical scenarios. These limit the possibility of obtaining behavioural data, rely on memory and imagination, and hinder the investigation of mechanisms, and I indicated that this can be addressed through experiments such as the Walking the plank paradigm. Furthermore, I indicated that self-reports tended to poorly correspond to observed behaviour with one-in-five participants misreporting on safety voice. This is important for the safety voice literature because it underscores the need to i) move beyond the reliance on post-hoc and report-based methodologies and ii) clarify how findings from report-based measures can be generalised to behaviour with more certainty. For instance, whilst the critical incident interview study by Aydon and colleagues (2016) is amongst the most comprehensive publications within the literature, and one of my favourite readings during this thesis, because of these methodological limitations the identified antecedents (e.g., supportive role models, low workload, learning opportunities) would benefit from direct behavioural observation in controlled environments.

To this end, I made tractable new research by providing a manual for the Walking the Plank paradigm (see Appendix C), and additional files will be made available as supplementary materials (i.e., the safety voice dictionaries and the CVR dataset) upon publication of the submitted articles[34]. For instance, and in addition to the suggested directions in Chapter 4, scholars may want to use these materials to triangulate new evidence across methodologies. For example, I would encourage future research to triangulate other datasets with the CVR dataset, apply the safety voice dictionaries to investigate how variables of interest (e.g., power distance) shape speech, or to explore to what extent available safety voice questionnaire items (Tucker et al., 2008) capture observed speech.

Finally, the proposed Walking the Plank paradigm may be adapted for investigating broader safety-related phenomena. For instance, because the paradigm enables the observation of psychological phenomena during an apparent hazard, it can be used for exploring hypotheses on the extent to which people have situational awareness about hazards (Stanton et al., 2001), rely on others for information about hazardous situations (e.g., the bystander effect; P. Fischer et al., 2011), are obedient when authority figures fall victim to hazards (i.e., extending research on obedience to authority; Milgram, 1974), engage in safety citizenship behaviours (Didla et

---

[34] These materials can also be provided upon request.

al., 2009), or are willing to self-sacrifice to avoid harm to others by walking the plank oneself[35]. In particular, the Walking the Plank paradigm might be used in an experimental investigation of the Human Factors Analysis and Classification System (HFACS; Wiegmann & Shappell, 2016). For instance, and whilst it appears unlikely that experiments can fully simulate complex safety management systems, the paradigm may be adapted to investigate the interactions between organisational influences (e.g., by providing a safety training and outlining safety accountabilities beforehand), inadequate supervision (e.g., by confederates failing to correct hazards) and unsafe preconditions (e.g., by assessing adverse mental states). I encourage future research to adapt the provided protocol of the Walking the Plank paradigm (see Appendix C) to explore these options in research on safety voice and safety behaviours more broadly.

Implications for preventing harm

Insights on conceptualising and assessing the behavioural nature of safety voice have implications for researchers and practitioners aiming to prevent harm, and lessons can be learned for designing interventions.

Interventions need to improve the effectiveness of safety voice

First, research has assumed that safety voice emerges from safety threats and is essential for avoiding accidents (Bienefeld & Grote, 2012; Enomoto & Geisler, 2017; Gladwell, 2008; Noort et al., 2019b; Soeters & Boer, 2000; Tucker et al., 2008). Yet, my evidence only supported a loose contingency upon risk and the need for good safety listening. That is, I indicated that whilst risk from hazards (perceived and actual) predicts safety voice, safety voice

[35] In an early pilot that involved a version of the paradigm where potential victims sat on the plank (discussed in Chapter 2), I observed two cases in which participants offered to sit on the plank instead of the research assistant. Whilst this happened infrequently and it might have been that participants felt they could mitigate risk by not fully sitting on the plank, this means the Walking the Plank paradigm enables research on heroism and self-sacrifice.

requires that people need to be aware and concerned about hazards (see Chapter 5) and, moreover, aviation accidents still occurred when people spoke-up to a high degree (see Chapter 6) because safety listening was often inadequate. Furthermore, despite risks during aviation accidents being extreme, variation persisted in the extent to which flight crew spoke-up. This means that the literature cannot assume that safety voice is sufficient for avoiding accidents (i.e., social factors inhibit the mitigation of accidents), or that accidents occur because people fail to speak-up (i.e., voice and accidents co-occurred). This supports a proposition by Jones and Kelly (2014) that listening is an important factor in avoiding accidents, which I conceptualised as 'safety listening', and contradicts a central assumption in the literature (i.e., that interventions are best targeted at making safety voice more likely). Essentially, my findings underscore that safety management interventions need to make safety voice more likely *and* effective by enabling people to be concerned about safety and engage in safety voice and safety listening.

Interventions need to be tailored to hazardous scenarios

Second, the thesis indicates that safety voice interventions are optimally designed by being tailored towards the situation. That is, the interventions investigated in this thesis (i.e., hazard salience, responsibility, encouragements, CRM; Barzallo Salazar et al., 2014; Burris, 2012; Duan et al., 2017; Helmreich et al., 1990; Kanki et al., 2019; Tucker et al., 2008) varied in their effectiveness based on (perceived) properties of the hazardous situation. Interventions were more effective when participants engaged in knowledge-based safety voice and held stronger safety concerns. This means that interventions should be designed according to the level of risk, with low levels of risk requiring different interventions (i.e., to increase the extent to which people are concerned) from high degrees of risk (i.e., to remove inhibitors to safety voice and safety listening). Whilst safety voice research has not distinguished between levels of risk in the design of interventions, this has long been a central notion in how risk research informs safety management (Renn, 1998), and applying this to safety voice may enable better interventions through a solid grounding in the properties of hazards that people encounter.

In addition, intervention success varied according to the stage of the hazard (see Chapter 5) and cultural norms. For instance, in Chapter 6, I indicated that the introduction of CRM training has led to historic changes in safety voice and safety listening. Yet, and indicating the importance of addressing the situated nature of safety voice, safety voice and safety listening only improved for airlines registered in low power distance countries. This suggests that the situated nature of safety voice reduces the effectiveness of generic interventions, and may explain why safety voice interventions (O'Donovan & McAuliffe, 2020) and antecedents (see Chapter 3) have had mixed success. This is particularly important because studies rarely establish the extent to which scenarios elicit safety concerns (cf. Schwappach & Gehring, 2014c), the aspects of safety voice that are altered by interventions, the timing of changes in safety voice (cf. Krenz et al., 2020), how others listen to safety voice during real accidents, or the impact of national culture on interventions (see Chapter 3 and 6). Thus, I expand safety voice theory by indicating that previously indicated antecedents (Chapter 3) may have nuanced in-situ effects on safety voice behaviour (e.g., shared safety knowledge may only improve the sharing of information, not motivations to prevent harm; Phelps & Reed, 2016).

Finally, this contributes to the safety management literature by indicating that after thirty years a need still exists (Helmreich et al., 1990) to improve the training effectiveness of programs that aim to improve communication during safety-critical scenarios (Kanki et al., 2019; King et al., 2008; Omura et al., 2017) in high power distance countries. Furthermore, this also suggests that the behaviour-based safety interventions reviewed by Tuncel and colleagues (2006; though without reference to national culture) may benefit from being tailored to national cultural environments.

Limitations

Limitations were discussed within the presented articles, and for the purpose of conciseness here I only discuss limitations of general importance for conceptualising, assessing and intervening on safety voice.

First, the scope of this research was limited by the choices made. By focussing on safety voice as an individual behaviour rather than collective action (Wilkinson et al., 2020), or shared perceptions on behavioural trends within groups (e.g., voice climate; Ditchburn & Hames, 2014), the systematic review and presented studies did not capture safety voice that does not involve acts of interpersonal communication. This choice enabled targeted studies (e.g., enabling fewer variables and participants), provided a manageable number of articles to be systematically reviewed and provided insights specific to the behavioural nature of safety voice. However, it means that further research is needed to evaluate how findings on safety voice can be generalised beyond interpersonal communication (e.g., to warning symbols, or press conferences during times of crisis; Matthews et al., 2014; Mullin, 2003).

Furthermore, to reiterate, the systematic literature review was limited to the articles retrieved by the search strategy and inclusion criteria, and relevant articles may not have been included when these were not uncovered by the search terms. Where I became aware of relevant articles not covered by the systematic literature review (e.g., Weiss et al., 2014, 2017, 2018) or when key articles were published during the duration of the thesis (Krenz et al., 2020, e.g., 2019; Peadon et al., 2020), I aimed for a complete account by integrating this work in other sections of this thesis. In addition, by emphasising the act of speaking-up about safety, research contained in other literatures such as protective action decision-making (Lindell & Perry, 2012) or risk and crisis communication (Reynolds & Seeger, 2005) may have been missed. I aimed to mitigate this by systematically reviewing research on safety voice and reviewing relevant literature within other sections of this thesis (e.g., the protective action decision-model was cited in Chapter 5; crisis communication was referred to in Chapter 6). However, the extent to which other literatures (beyond voice, safety and teamworking, see Chapter 3) draw on similar ideas as safety voice remains an empirical question for future research.

Second, I aimed for high internal validity for the presented studies, but additional research is needed to establish whether findings replicate. In particular, I established that the Walking the Plank paradigm can address challenges to obtaining data on safety voice behaviour, but the safety concern measure in these studies involved a single, self-reported item that was completed after the scenario. Single item questionnaires can be successful (Konrath et al., 2018; Robins et al., 2001) and I agreed with others (P. Fischer et al., 2006) that post-scenario procedures are necessary to maintain successful deception. Moreover, I aimed to address this issue by providing a safety concern dictionary that indicated a strong relationship to safety voice speech. Yet, this measure was developed and evaluated within the same dataset (i.e., potentially providing common methods bias; Podsakoff & Organ, 1986) and I encourage future research intending to adopt the report-based measure to increase the number of items. In addition, future research may utilise the provided CVR dataset to reveal whether the indicated good reliability holds beyond the modest sample of the data, and develop intersubjective measures. Whilst my results indicate safety voice is a sensemaking process (e.g., people voiced less when others engaged in poor safety listening), analyses for Chapter 5 captured individuals' utterances and future research should investigate how safety concerns can be tracked in conversation (for a method, see: Heasman & Gillespie, 2019) because sensemaking is intersubjective (Gillespie & Cornish, 2010).

Third, and of special importance to internal validity, it was not always clear to what extent studies established safety silence. Because safety concerns are rarely assessed (see Chapter 4), the systematic review could not establish whether less safety voice involved concerned or unconcerned silence. This means that the systematic literature review provides no clear conclusions on safety silence. This is a limitation of the literature and, to reiterate, this indicates that safety voice studies need to establish the extent to which people are concerned in order to establish safety silence (when this is of interest to researchers). The analysis of the CVR transcripts illustrates this: in the study I could only draw conclusions on safety voice and safety listening, but not on safety silence, because I could not assess the extent to which the safety concern dictionary could be validly applied to the dataset. That is, the interpretation of social-psychological phenomena beyond a text itself (e.g., cognitive states such as safety concerns) is limited because it involves the interpretation of speakers' intended meaning (for a discussion of methods and assumptions underpinning the analysis of speech, see: Hammersley, 2003). The limited number of utterances for individual pilots requires a better understanding of how pilots understood their context (Gillespie & Cornish, 2010) than the historic nature of the data permitted (i.e., new data could not be obtained to triangulate other safety concern measures). It may be fruitful to explore the extent to which safety concerns can be inferred form high levels of safety voice or situational characteristics because safety silence can be scaled based on the extent of safety voice speech. Indeed, safety voice and concern dictionaries may be tailored to specific contexts, and future research should establish the validity of the dictionaries in new contexts by ascertaining the extent to which participants in the given context are concerned. However, arguably, researchers should provide clarity by only adopting the term 'safety silence' when concerns can be ascertained.

Fourth, in terms of external validity, I suggested that additional research is needed to establish the extent to which findings from the Walking the plank paradigm generalise to other contexts. To reiterate, the external validity of experiments is debated (Jiménez-Buedo & Miller, 2010) and I indicated that the findings from the Walking the plank paradigm can generalise: the findings on poor safety listening (i.e., worry and discouragements; Chapters 4 and 5) were reflected in the unfavourable effect of high power distance and safety listening in natural environments (Chapter 6). Furthermore, the behaviour-report gap indicated in Chapter 4 shows that conclusions based on direct observations may be generalised to report-based measures, and vice versa, when inaccurate self-reports are accounted for.

Finally, additional research is needed to provide empirical evidence for propositions that remain conceptual. Above, I derived theoretical insights and whilst I indicated how my evidence supports these propositions, more data is needed on the Threat Mitigation Model. Furthermore, whilst my findings have clear implications for threat mitigation because safety voice outcomes can include the prevention of harm (Blanco et al., 2009; Seiden et al., 2006; Tucker & Turner, 2015), I did not present novel data on the extent to which safety voice prevents harm. This is because the Walking the Plank paradigm did not contain actual risk and the CVR study did not present an appropriate control group for normal flights or near-misses. Yet, future research may investigate how the prevention of harm varies according to the nuanced manifestation of safety concerns, safety voice, and safety listening.

Concluding remarks

By presenting four articles, in this thesis I contributed insights into how safety voice should be optimally conceptualised, assessed and intervened on. Insights were contributed towards conceptualising safety voice as a unique concept for understanding the extent to which safety critical information flows effectively during hazardous scenarios (Westrum, 2014), with the situated nature of safety voice supporting its incorporation into sociotechnical models (e.g., the Swiss Cheese model; Perrow, 2011; Reason, 1990; Vaughan, 1999) because the social and technical properties of hazardous scenarios can prevent accidents when they elicit safety concerns, safety voice and safety listening.

In this thesis, I addressed a need for better understanding the behavioural nature of safety voice, and I showed that safety voice has a unique scope compared to research on safety management and employee voice through its narrow emphasis on the communication of safety knowledge and motivations during hazardous scenarios, and broad context of application. Safety voice and safety silence, though distinct acts, can be captured within the same theoretical models because safety silence can be scaled based upon the extent to which concerned individuals speak-up during hazardous situations. In particular, the acts of safety voice and silence are manifested as highly nuanced speech that are elicited by, and reflect, the technical (i.e., perceived and actual risk, progression of time) and social characteristics (i.e., norms for communication) of encountered hazards. Moreover, understanding the interaction between social and technical factors is essential for understanding how safety voice may prevent accidents, because safety voice nearly always occurs when levels of real risk are high. Yet, the extent of safety voice engagement varies according to social properties of hazardous scenarios such as safety listening and norms for communication. Thus, I revealed the behavioural nature of safety voice to be highly ecological (i.e., with antecedents, the act of safety voice and outcomes being identifiable across levels of analysis) and situated (i.e., with safety threats eliciting degrees of safety concerns, safety voice and safety listening), and proposed the Threat Mitigation Model of safety voice. This enables better safety management by providing a specific concept for the interpersonal communication of concerned individuals within sociotechnical systems (Appelbaum, 1997; Reason, 2000) that emphasises that failures in safety management can emerge when sociotechnical systems poorly enable people to be concerned, to voice and to listen.

Because of this ecological and situated nature of safety voice, the assessment of safety voice and interventions to reduce safety silence, should be grounded in direct behavioural observations that establish the extent to which people engage in safety voice and safety listening, and ascertain the extent to which hazardous scenarios pose actual levels of risk or elicit risk perceptions. The literature has relied on methodologies that provide valuable insights, but rarely directly establish safety voice behaviours, or safety concerns, during hazardous scenarios. I identified nine methodological shortfalls of the literature and made tractable the investigation of new hypotheses by providing and illustrating a novel experimental paradigm (i.e., Walking the Plank) and dataset containing actual, field-based, safety voice behaviours (i.e., the CVR dataset). Moreover, and underscoring the situated nature of safety voice, I indicated that interventions for reducing safety silence require that people are concerned. Yet, interventions should not only improve the likelihood of safety voice, but improve its effectiveness (e.g., by improving safety listening) because accidents can still occur when safety-critical staff are concerned and speak-up.

Thus, in this thesis, I revealed that safety voice enables a better understanding of accident causation than broader safety management concepts (e.g., safety culture, safety citizenship, safety leadership) by providing new insights into the extent to which people engage in interpersonal communication about safety within sociotechnical systems: safety voice behaviour is rooted in the social and technical properties of hazardous situations that impact on the extent to which people are concerned, voice and listen, with all being required to enable accident mitigation. Future research should design studies and interventions that move beyond antecedents that increase the likelihood of safety voice and establish how hazardous scenarios enable individuals to be concerned about safety, engage in a higher degree of nuanced speech about safety themes, and listen effectively when concerns are raised. Thus, the investigation of safety voice may become more nuanced and contextual, informative for how accidents emerge, and effective in designing interventions aimed at improving communication to mitigate safety threats. In particular, establishing and addressing the social and technical properties of hazardous scenarios is essential for preventing accidents by enabling individuals to listen and speak-up about safety concerns.

BIBLIOGRAPHY

Aarts, H., Custers, R., & Marien, H. (2008). Preparing and Motivating Behavior Outside of Awareness. *Science*, *319*(5870), 1639–1639. https://doi.org/10.1126/science.1150432

Alingh, C. W., Van Wijngaarden, J. D. H., Van De Voorde, K., Paauwe, J., & Huijsman, R. (2019). Speaking up about patient safety concerns: The influence of safety management approaches and climate on nurses' willingness to speak up. *BMJ Quality and Safety*, *28*(1), 39–48. https://doi.org/10.1136/bmjqs-2017-007163

Almeida, F., Superior, I., Gaya, P., Queirós, A., & Faria, D. (2017). Strengths and Limitations of Qualitative and Quantitative Research Methods. *European Journal of Education Studies*, *3*(9), 369–387. https://doi.org/10.5281/zenodo.887089

Anicich, E. M., Swaab, R. I., & Galinsky, A. D. (2015). Hierarchical cultural values predict success and mortality in high-stakes teams. *PNAS*, *112*(5), 1338–1343. https://doi.org/10.1073/pnas.1408800112

Antonsen, S. (2009). Safety culture and the issue of power. *Safety Science*, *47*(2), 183–191. https://doi.org/10.1016/j.ssci.2008.02.004

APA. (1959). Ethical standards of psychologists. *American Psychologist*, *14*, 279–282. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=ovfta&NEWS=N&AN=00000487-195906000-00004.

APA. (2017). Ethical Principles of Psychologists and Code of Conduct. In *Psychotherapy*. https://www.apa.org/ethics/code/ethics-code-2017.pdf

Appelbaum, S. H. (1997). Socio-technical systems theory: an intervention strategy for organizational development. *Management Decision*, *35*(6), 452–463. https://doi.org/10.1108/00251749710173823

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*. https://doi.org/10.1037/h0093718

Aubin, D., & King, S. (2015). Developing a culture of safety: Exploring students? perceptions of errors in an interprofessional setting. *Journal of Interprofessional Care*, *29*(6), 646–648. https://doi.org/10.3109/13561820.2015.1045060

Aviation-Safety Network. (2019a). *Statistics by period*. https://aviation-safety.net/statistics/period/stats.php

Aviation-Safety Network. (2019b). *Transcripts*. aviation-safety.net/investigation/cvr/transcripts/

Aydon, L., Hauck, Y., Zimmer, M., & Murdoch, J. (2016). Factors influencing a nurse's decision to question medication administration in a neonatal clinical care unit. *Journal of Clinical Nursing*, *25*(17–18), 2468–

2477. https://doi.org/10.1111/jocn.13277

Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing Construct Validity in Organizational Research. *Administrative Science Quarterly*, *36*(3), 421–458. http://www.jstor.org/stable/2393203

Barling, J., Loughlin, C., & Kelloway, E. K. (2002). Development and test of a model linking safety-specific transformational leadership and occupational safety. *Journal of Applied Psychology*, *87*(3), 488–496. https://doi.org/10.1037/0021-9010.87.3.488

Barlow, M., Watson, B., & Rudolph, J. (2019). How being a great Receiver can change the game in speaking up conversations. *MedEdPublish*, *8*(3), 1–6. https://doi.org/10.15694/mep.2019.000204.1

Barnett, T. (1992). A preliminary investigation of the relationship between selected organizational characteristics and external whistleblowing by employees. *Journal of Business Ethics*, *11*(12), 949–959. https://doi.org/10.1007/BF00871961

Bartlett, F. C. (1932). Remembering: A Study in Experimental and Social Psychology. *Cambridge, Social Psychology*, 1–11. https://doi.org/10.1111/j.2044-8279.1933.tb02913.x

Barton, M. A., & Sutcliffe, K. M. (2009). Overcoming dysfunctional momentum: Organizational safety as a social achievement. *Human Relations*, *62*(9), 1327–1356. https://doi.org/10.1177/0018726709334491

Barzallo Salazar, M. J., Minkoff, H., Bayya, J., Gillett, B., Onoriode, H., Weedon, J., Altshuler, L., & Fisher, N. (2014). Influence of surgeon behavior on trainee willingness to speak up: A randomized controlled trial. *Journal of the American College of Surgeons*, *219*(5), 1001–1007. https://doi.org/10.1016/j.jamcollsurg.2014.07.933

Bashshur, M. R., & Oc, B. (2014). When Voice Matters: A Multilevel Review of the Impact of Voice in Organizations. *Journal of Management*, *41*(5), 1530–1554. https://doi.org/10.1177/0149206314558302

Batson, C. D., Eklund, J. H., Chermok, V. L., Hoyt, J. L., & Ortiz, B. G. (2007). An Additional Antecedent of Empathic Concern: Valuing the Welfare of the Person in Need. *Journal of Personality and Social Psychology*, *93*(1), 65–74. https://doi.org/10.1037/0022-3514.93.1.65

Bauer, M. W., & Gaskell, G. (1999). Towards a Paradigm for Research on Social Representations. *Journal for the Theory of Social Behaviour*, *29*(2), 163–186. https://doi.org/10.1111/1468-5914.00096

Baumeister, R. F. (1982). A self-presentational view of social phenomena. *Psychological Bulletin*, *91*(1), 3–26. https://doi.org/10.1037/0033-2909.91.1.3

BBC. (2015). *How speaking up can save lives*. http://www.bbc.co.uk/news/health-33544778

Bell, E., & Bryman, A. (2007). The ethics of management research: An exploratory content analysis. *British*

Journal of Management, *18*(1), 63–77. https://doi.org/10.1111/j.1467-8551.2006.00487.x

Belyansky, I., Martin, T. R., Prabhu, A. S., Tsirline, V. B., Howley, L. D., Phillips, R., Sindram, D., Heniford, B. T., & Stefanidis, D. (2011). Poor Resident-Attending Intraoperative Communication May Compromise Patient Safety. *Journal of Surgical Research*, *171*(2), 386–394. https://doi.org/10.1016/j.jss.2011.04.011

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, *74*(3), 183–200. https://doi.org/10.1037/h0025146

Bennett, S., Banyard, V. L., & Garnhart, L. (2014). To Act or Not to Act, That Is the Question? Barriers and Facilitators of Bystander Intervention. *Journal of Interpersonal Violence*, *29*(3), 476–496. https://doi.org/10.1177/0886260513505210

Berger, P. L., & Luckmann, T. (1966). The Social Construction of Reality. In *Penguin Group*. Penguin Books.

Beus, J. M., McCord, M. A., & Zohar, D. (2016). Workplace safety: A review and research synthesis. *Organizational Psychology Review*, *6*(4), 352–381. https://doi.org/10.1177/2041386615626243

Bickhoff, L., Levett-jones, T., & Sinclair, P. M. (2016). Nurse Education Today Rocking the boat - nursing students' stories of moral courage: A qualitative descriptive study. *Nurse Education Today*, *42*, 35–40. https://doi.org/10.1016/j.nedt.2016.03.030

Bienefeld, N., & Grote, G. (2012). Silence that may kill: When aircrew members don't speak up and why. *Aviation Psychology and Applied Human Factors*, *2*(1), 1–10. https://doi.org/10.1027/2192-0923/a000021

Bienefeld, N., & Grote, G. (2014). Speaking up in ad hoc multiteam systems: Individual-level effects of psychological safety, status, and leadership within and across teams. *European Journal of Work and Organizational Psychology*, *23*(6), 930–945. https://doi.org/10.1080/1359432X.2013.808398

Bigman, C. A., Mello, S., Sanders-Jackson, A., & Tan, A. S. L. (2019). Speaking up about Lighting up in Public: Examining Psychosocial Correlates of Smoking and Vaping Assertive Communication Intentions among U.S. Adults. *Health Communication*, *34*(4), 500–510. https://doi.org/10.1080/10410236.2018.1428849

Blanco, M., Clarke, J. R., Martindell, D., & Denise, M. (2009). Wrong site surgery. *AORN Journal*, *90*(7), 814–816. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1857133/%5Cnhttp://www.ncbi.nlm.nih.gov/pmc/articles/PMC1857133/pdf/814A.pdf

Blass, T. (1999). The Milgram Paradigm After 35 Years: Some Things We Now Know About Obedience to Authority1. *Journal of Applied Social Psychology*, *29*(5), 955–978. https://doi.org/10.1111/j.1559-1816.1999.tb00134.x

Bolsin, S. N., Faunce, T., & Oakley, J. (2005). Practical virtue ethics: Healthcare whistleblowing and portable digital technology. *Journal of Medical Ethics*, *31*(10), 612–618. https://doi.org/10.1136/jme.2004.010603

Botero, I. C., & Van Dyne, L. (2009). Employee Voice Behavior: Interactive Effects of LMX and Power Distance in the United States and Colombia. *Management Communication Quarterly*, *23*(1), 84–104. https://doi.org/10.1177/0893318909335415

Boyce, A. S., Nieminen, L. R. G., Gillespie, M. A., Ryan, A. M., & Denison, D. R. (2015). Which comes first, organizational culture or performance? A longitudinal study of causal priority with automobile dealerships. *Journal of Organizational Behavior*, *36*(3), 339–359. https://doi.org/10.1002/job.1985

BPS. (2014). Code of Human Research Ethics. In *The Briish Psychological Society Guidelines*. https://www.bps.org.uk/sites/www.bps.org.uk/files/Policy/Policy - Files/BPS Code of Human Research Ethics.pdf

BPS. (2018). *Code of Ethics and Conduct* (Issue February). https://www.bps.org.uk/sites/www.bps.org.uk/files/Policy/Policy - Files/BPS Code of Ethics and Conduct %28Updated July 2018%29.pdf

Bradley, B. (2012). Doing Away with Harm. *Philosophy and Phenomenological Research*, *85*(2), 390–412. https://www-jstor-org.gate3.library.lse.ac.uk/stable/41721239%0A

Brashers, D. E. (2001). Communication and uncertainty management. *Journal of Communication*, *51*(3), 477–497. https://doi.org/10.1093/joc/51.3.477

Brinsfield, C. T. (2013). Employee silence motives: Investigation of dimensionality and development of measures. *Journal of Organizational Behavior*, *34*(5), 671–697. https://doi.org/10.1002/job.1829

Brinsfield, C. T., Edwards, M. S., & Greenberg, J. (2009). Voice and silence in organizations: Historical review and current conceptualizations. In J. Greenberg & M. S. Edwards (Eds.), *Voice and Silence in Organizations*. Emerald Group Publishing.

Bromiley, M., & Mitchell, L. (2009). Would you Speak up if the Consultant Got it Wrong? …And Would you Listen if Someone Said You'd Got it Wrong? *Journal of Perioperative Practice*, *19*(10), 326–329. https://doi.org/10.1177/175045890901901004

Bronfenbrenner, U. (1977). Toward an Experimental Ecology of Human Development. *American Psychologist*, *32*(7), 513–531. https://doi.org/10.1037/0003-066X.32.7.513

Brown, C. A., & Lilford, R. J. (2006). The stepped wedge trial design: A systematic review. *BMC Medical Research Methodology*, *6*, 1–9. https://doi.org/10.1186/1471-2288-6-54

Brown, J. S., Laundre, J. W., & Gurung, M. (1999). The Ecology of Fear: Optimal Foraging, Game Theory, and Trophic Interactions. *Journal of Mammalogy*, *80*(2), 385–399. https://doi.org/10.2307/1383287

Bruers, S., & Braeckman, J. (2014). A Review and Systematization of the Trolley Problem. *Philosophia (United States)*, *42*(2), 251–269. https://doi.org/10.1007/s11406-013-9507-5

Burke, B. L., Martens, A., & Faucher, E. H. (2010). Two decades of terror management theory: A meta-analysis of mortality salience research. *Personality and Social Psychology Review*, *14*(2), 155–195. https://doi.org/10.1177/1088868309352321

Burris, E. R. (2012). The risks and rewards of speaking up: Managerial response to employee voice. *Academy of Management Executive*, *55*(4), 851–875. https://doi.org/10.5465/amj.2010.0562

Burris, E. R., Detert, J. R., & Romney, A. C. (2013). Speaking up vs. being heard: The disagreement around and outcomes of employee voice. *Organization Science*, *24*(1), 22–38. https://doi.org/10.1287/orsc.1110.0732

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology*, *67*(2), 319–333. https://doi.org/10.1037/0022-3514.67.2.319

Chassang, G. (2017). The impact of the EU general data protection regulation on scientific research. *Ecancer Medical Science*, *11*, 1–12. https://doi.org/10.3332/ecancer.2017.709

Chou, S. Y., & Chang, T. (2017). Employee Silence and Silence Antecedents: A Theoretical Classification. *International Journal of Business Communication*, *00*(0), 1–26. https://doi.org/10.1177/2329488417703301

Christensen-Szalanski, J. J. J., & Willham, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes*, *48*(1), 147–168. https://doi.org/10.1016/0749-5978(91)90010-Q

Christian, M. S., Bradley, J. C., Wallace, J. C., & Burke, M. J. (2009). Workplace safety: A meta-analysis of the roles of person and situation factors. *Journal of Applied Psychology*, *94*(5), 1103–1127. https://doi.org/10.1037/a0016172

Christie, T., Groarke, L., & Sweet, W. (2008). Virtue ethics as an alternative to deontological and consequential reasoning in the harm reduction debate. *International Journal of Drug Policy*, *19*(1), 52–58. https://doi.org/10.1016/j.drugpo.2007.11.020

Cocklin, J. T. (2004). Swissair 111 Human Factors: Checklist and Cockpit Communications. *Journal of Air Transportation*, *9*(3), 19–42. https://search.proquest.com/docview/232853869?accountid=9630

Conchie, S. M., Taylor, P. J., & Donald, I. J. (2012). Promoting safety voice with safety-specific transformational leadership: The mediating role of two dimensions of trust. *Journal of Occupational Health Psychology*, *17*(1), 105–115. https://doi.org/10.1037/a0025101

Cookson, S. (2015). *‘ We Need Priority Please ’ Mitigated Speech in the Crash of Avianca Flight 052*. 578–583.

Corti, K. C. (2015). *Developing the cyranoid method of mediated interpersonal communication in a social psychological context: Applications in person peception, human-computer interaction, and first-person research*. London School of Economics and Political Science.

Cox, D. R., & Oakes, D. (1984). *Analysis of survival data*. Chapman & Hall.

Crockett, M. J., Kurth-Nelson, Z., Siegel, J. Z., Dayan, P., & Dolan, R. J. (2014). Harm to others outweighs harm to self in moral decision making. *Proceedings of the National Academy of Sciences*, *111*(48), 17320–17325. https://doi.org/10.1073/pnas.1408988111

Custers, R., & Aarts, H. (2010). The Unconscious Will: How the Pursuit of Goals Operates Outside of Conscious Awareness. *Science*, *329*(5987), 47–50. https://doi.org/10.1126/science.1188595

Daniels, M. A., & Greguras, G. J. (2014). Exploring the Nature of Power Distance: Implications for Micro- and Macro-Level Theories, Processes, and Outcomes. *Journal of Management*, *40*(5), 1202–1229. https://doi.org/10.1177/0149206314527131

Darley, J. M., & Latane, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, *8*(4, Pt.1), 377–383. https://doi.org/10.1037/h0025589

Davis, R. E., Vincent, C., Henley, A., & McGregor, A. (2013). Exploring the care experience of patients undergoing spinal surgery: a qualitative study. *Journal of Evaluation in Clinical Practice*, *19*(1), 132–138. https://doi.org/10.1111/j.1365-2753.2011.01783.x

de Ruiter, J. P., & Albert, S. (2017). An Appeal for a Methodological Fusion of Conversation Analysis and Experimental Psychology. *Research on Language and Social Interaction*, *50*(1), 90–107. https://doi.org/10.1080/08351813.2017.1262050

Del Boca, F. K., & Noll, J. A. (2000). Truth or consequences: the validity of self-report data in health services research on addictions. *Addiction*, *95*(11), 347–360. https://doi.org/10.1080/09652140020004278

Delisle, M., Grymonpre, R., Whitley, R., & Wirtzfeld, D. (2016). Crucial Conversations: An interprofessional learning opportunity for senior healthcare students. *Journal of Interprofessional Care*, *30*(6), 777–786. https://doi.org/10.1080/13561820.2016.1215971

Detert, J. R. (2003). To speak or not to speak: Multi-level leadership influences on organizational voice. In *Dissertation Abstracts International Section A: Humanities and Social Sciences* (Vol. 64, Issues 5-A). http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc4&NEWS=N&AN=2003-95021-

116

Detert, J. R., & Burris, E. R. (2007). Leadership Behavior and Employee Voice: Is the Door Really Open? *Academy of Management Journal*, *50*(4), 869–884. https://doi.org/10.5465/amj.2007.26279183

Detert, J. R., & Treviño, L. K. (2010). Speaking Up to Higher-Ups: How Supervisors and Skip-Level Leaders Influence Employee Voice. *Organization Science*, *21*(1), 249–270. https://doi.org/10.1287/orsc.1080.0405

Didla, S., Mearns, K., & Flin, R. (2009). Safety citizenship behaviour: a proactive approach to risk management. *Journal of Risk Research*, *12*(3–4), 475–483. https://doi.org/10.1080/13669870903041433

Ditchburn, G. J., & Hames, K. (2014). Voice Climate Perceptions: A Multidimensional Model as a Determinant of Affective Commitment, Work Engagement, Neglect and Exit. *Australasian Journal of Organisational Psychology*, *7*(e3), 1–9. https://doi.org/10.1017/orp.2014.3

Dixon-Woods, M., Cavers, D., Agarwal, S., Annandale, E., Arthur, A., Harvey, J., Hsu, R., Katbamna, S., Olsen, R., Smith, L., Riley, R., & Sutton, A. J. (2006). Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. *BMC Medical Research Methodology*, *6*(1), 35–48. https://doi.org/10.1186/1471-2288-6-35

Douglas, M. (1992). *Risk and blame: essays in cultural theory*. Routledge.

Driscoll, G. (2002). *Cockpit conversation: A communication analysis of three aviation accidents*. University of Colorado.

Duan, J., Li, C., Xu, Y., & Wu, C. H. (2017). Transformational leadership and employee voice behavior: A Pygmalion mechanism. *Journal of Organizational Behavior*, *38*(5), 650–670. https://doi.org/10.1002/job.2157

Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, *6*(May), 1–11. https://doi.org/10.3389/fpsyg.2015.00621

Edmondson, A. C. (1999). Psychological Safety and Learning Behavior in Work Teams. *Administrative Science Quarterly*, *44*(2), 350–383. https://doi.org/10.2307/2666999

Edmondson, A. C. (2003). Speaking Up in the Operating Room: How Team Leaders Promote Learning in Interdisciplinary Action Teams. *Journal of Management Studies*, *40*(6), 1419–1452. https://doi.org/10.1111/1467-6486.00386

Edmondson, A. C., & Lei, Z. (2014). Psychological safety: The history, renaissance, and future of an interpersonal construct. *Annual Review of Organizational Psychology and Organizational Behavior*, *1*, 23–43. https://doi.org/10.1146/annurev-orgpsych-031413-091305

Enomoto, C. E., & Geisler, K. R. (2017). Culture and plane crashes: A cross-country test of the gladwell hypothesis. *Economics and Sociology*, *10*(3), 281–293. https://doi.org/10.14254/2071-789X.2017/10-3/20

Entwistle, V. a, McCaughan, D., Watt, I. S., Birks, Y., Hall, J., Peat, M., Williams, B., & Wright, J. (2010). Speaking up about safety concerns: multi-setting qualitative study of patients' views and experiences. *BMJ Quality & Safety*, *19*(6), e33–e33. https://doi.org/10.1136/qshc.2009.039743

Epstein, B. (2018). *Social ontology*. https://plato.stanford.edu/archives/sum2018/entries/social-ontology/

Erez, M., & Gati, E. (2004). A dynamic, multi-level model of culture: From the micro level of the individual to the macro level of a global culture. *Applied Psychology: An International Review*, *53*(4), 583–598. https://doi.org/10.1111/j.1464-0597.2004.00190.x

Farh, C. I. C., & Chen, G. (2018). Leadership and member voice in action teams: Test of a dynamic phase model. *Journal of Applied Psychology*, *103*(1), 97–110. https://doi.org/10.1037/apl0000256

Farrell, D. (1983). Exit, Voice, Loyalty, and Neglect as Responses to Job Dissatisfaction: A Multidimensional Scaling Study. *Academy of Management Journal*, *26*(4), 596–607. https://doi.org/10.2307/255909

Fay, B. (1996). *Contemporary philosophy of social science: a multicultural approach*. Blackwell Publishing Ltd.

Fioratou, E., Flin, R., Glavin, R., & Patey, R. (2010). Beyond monitoring: Distributed situation awareness in anaesthesia. *British Journal of Anaesthesia*, *105*(1), 83–90. https://doi.org/10.1093/bja/aeq137

Fischer, G. W., Morgan, M. G., Fischhoff, B., Nair, I., & Lave, L. B. (1991). What Risks Are People Concerned About. *Risk Analysis*, *11*(2), 303–314. https://doi.org/10.1111/j.1539-6924.1991.tb00606.x

Fischer, P., Greitemeyer, T., Pollozek, F., & Frey, D. (2006). The unresponsive bystander: Are bystanders more responsive in dangerous emergencies? *European Journal of Social Psychology*, *36*(2), 267–278. https://doi.org/10.1002/ejsp.297

Fischer, P., Krueger, J. I., Greitemeyer, T., Vogrincic, C., Kastenmüller, A., Frey, D., Heene, M., Wicher, M., & Kainbacher, M. (2011). The bystander-effect: a meta-analytic review on bystander intervention in dangerous and non-dangerous emergencies. *Psychological Bulletin*, *137*(4), 517–537. https://doi.org/10.1037/a0023304

Fischer, U., & Orasanu, J. (2000). Error-Challenging Strategies: Their Role in Preventing and Correcting Errors. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *44*(1), 30–33. https://doi.org/10.1177/154193120004400109

Foushee, H. C. (1984). Dyads and triads at 35,000 feet: Factors affecting group process and aircrew performance. *American Psychologist*, *39*(8), 885–893. https://doi.org/10.1037/0003-066X.39.8.885

Foushee, H. C., & Manos, K. L. (1981). Information transfer within the cockpit: Problems in intracockpit communications. In *Information transfer problems in the aviation system (NASA TP-1875)*. NASA-Ames Research Center.

Francis, R. (2013). *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry Executive summary Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry: executive report*. https://doi.org/10.1016/j.ecns.2016.02.007

Francis, R. (2015). *Freedom to speak up: an independent review into creating an open and honest reporting culture in the NHS* (Issue February). https://freedomtospeakup.org.uk/wp-content/uploads/2014/07/F2SU_web.pdf

Franks, B., Bangerter, A., Bauer, M. W., Hall, M., & Noort, M. C. (2017). Beyond "Monologicality"? Exploring Conspiracist Worldviews. *Frontiers in Psychology*, *8*(1–16). https://doi.org/10.3389/fpsyg.2017.00861

Friedman, Z., Hayter, M. A., Everett, T. C., Matava, C. T., Noble, L. M. K., & Bould, M. D. (2015). Power and conflict: The effect of a superior's interpersonal behaviour on trainees' ability to challenge authority during a simulated airway emergency. *Anaesthesia*, *70*(10), 1119–1129. https://doi.org/10.1111/anae.13191

Frosch, D. L., May, S. G., Rendle, K. A. S., Tietbohl, C., & Elwyn, G. (2012). Authoritarian Physicians And Patients' Fear Of Being Labeled "Difficult" Among Key Obstacles To Shared Decision Making. *Health Affairs*, *31*(5), 1030–1038. https://doi.org/10.1377/hlthaff.2011.0576

Fuller, J. B., Marler, L. E., & Hester, K. (2006). Promoting felt responsibility for constructive change and proactive behavior: Exploring aspects of an elaborated model of work design. *Journal of Organizational Behavior*, *27*(8), 1089–1120. https://doi.org/10.1002/job.408

Gauld, R., & Horsburgh, S. (2014). Healthcare professional perspectives on quality and safety in New Zealand public hospitals: findings from a national survey. *Australian Health Review*, *38*(1), 109. https://doi.org/10.1071/AH13116

Geller, E. S., Roberts, D. S., & Gilmore, M. R. (1996). Predicting propensity to actively care for occupational safety. *Journal of Safety Research*, *27*(1), 1–8. https://doi.org/10.1016/0022-4375(95)00024-0

Gergen, K. J. (1973). Social psychology as history. *Journal of Personality and Social Psychology*, *26*(2), 309–320. https://doi.org/10.1037/h0034436

Gigerenzer, G. (1984). External Validity of Laboratory Experiments: The Frequency-Validity Relationship. *The American Journal of Psychology*, *97*(2), 185–195. https://doi.org/10.2307/1422594

Gillespie, A., & Cornish, F. (2010). Intersubjectivity: Towards a dialogical analysis. *Journal for the Theory of*

Social Behaviour, *40*(1), 19–46. https://doi.org/10.1111/j.1468-5914.2009.00419.x

Gillespie, A., & Reader, T. W. (2016). The Healthcare Complaints Analysis Tool: Development and reliability testing of a method for service monitoring and organisational learning. *BMJ Quality and Safety*, *25*(12), 937–946. https://doi.org/10.1136/bmjqs-2015-004596

Gkorezis, P., Panagiotou, M., & Theodorou, M. (2016). Workplace ostracism and employee silence in nursing: the mediating role of organizational identification. *Journal of Advanced Nursing*, *72*(10), 2381–2388. https://doi.org/10.1111/jan.12992

Gladwell, M. (2008). *Outliers: The story of success.* Little, Brown and Company.

Gleibs, I. H. (2014). Turning virtual public spaces into laboratories: Thoughts on conducting online field studies using social network sites. *Analyses of Social Issues and Public Policy*, *14*(1), 352–370. https://doi.org/10.1111/asap.12036

Gleibs, I. H. (2017). Are all "research fields" equal? Rethinking practice for the use of data from crowdsourcing market places. *Behavior Research Methods*, *49*(4), 1333–1342. https://doi.org/10.3758/s13428-016-0789-y

Goodwin, C. J. (2008). *Research in psychology methods and design, 5th edition*. John Wiley & Sons.

Griffin, M. A., & Neal, A. (2000). Perceptions of safety at work: A framework for linking safety climate to safety performance, knowledge, and motivation. *Journal of Occupational Health Psychology*, *5*(3), 347–358. https://doi.org/10.1037/1076-8998.5.3.347

Griffin, R. J., Dunwoody, S., & Neuwirth, K. (1999). Proposed model of the relationship of risk information seeking and processing to the development of preventive behaviors. *Environmental Research*, *80*(2 II). https://doi.org/10.1006/enrs.1998.3940

Gruman, J. A., & Saks, A. M. (2014). Being psychologically present when speaking up: employee voice engagement. In A. Wilkinson, J. Donaghey, T. Dundon, & R. B. Freeman (Eds.), *Handbook of Research on Employee Voice*. Edward Elgar Publishing.

Guldenmund, F. W. (2000). The nature of safety culture: a review of theory and research. *Safety Science*, *34*(1–3), 215–257. https://doi.org/10.1016/S0925-7535(00)00014-X

Gurung, G., Derrett, S., Gauld, R., & Hill, P. C. (2017). Why service users do not complain or have 'voice': a mixed-methods study from Nepal's rural primary health care system. *BMC Health Services Research*, *17*(1), 81. https://doi.org/10.1186/s12913-017-2034-5

Guyer, P., & Horstmann, R.-P. (2019). Idealism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 201). Metaphysics Research Lab, Stanford University.

https://plato.stanford.edu/archives/win2019/entries/idealism

Haavik, T. K. (2014). On the ontology of safety. *Safety Science*, *67*, 37–43. https://doi.org/10.1016/j.ssci.2013.09.004

Haavik, T. K., Antonsen, S., Rosness, R., & Hale, A. (2015). HRO and RE: A pragmatic perspective. *Safety Science*. https://doi.org/10.1016/j.ssci.2016.08.010

Habyarimana, J., & Jack, W. (2011). Heckle and Chide: Results of a randomized road safety intervention in Kenya. *Journal of Public Economics*, *95*(11–12), 1438–1446. https://doi.org/http://dx.doi.org/10.1016/j.jpubeco.2011.06.008

Hammersley, M. (2003). Conversation analysis and discourse analysis: Methods or paradigms? *Discourse and Society*, *14*(6), 751–781. https://doi.org/10.1177/09579265030146004

Hanson, D. (2017). Reducing Central Line-Associated Bloodstream Infection Rates in the Context of a Caring-Healing Environment. *Journal of Infusion Nursing*, *40*(2), 101–110. https://doi.org/10.1097/NAN.0000000000000212

Heale, R., & Forbes, D. (2013). Understanding triangulation in research. *Evidence Based Nursing*, *16*(4), 98–98. https://doi.org/10.1136/eb-2013-101494

Heasman, B., & Gillespie, A. (2019). Neurodivergent intersubjectivity: Distinctive features of how autistic people create shared understanding. *Autism*, *23*(4), 910–921. https://doi.org/10.1177/1362361318785172

Helmreich, R. L., Chidester, T., Foushee, H. C., Gregorich, S., & Wilhelm, J. A. (1990). How effective is cockpit resource management training? Exploring issues in evaluating the impact of programs to enhance crew coordination. *Flight Safety Digest*, *9*(5), 1–17. https://flightsafety.org/fsd/fsd_may90.pdf

Helmreich, R. L., Merritt, A. C., & Wilhelm, J. A. (1999). The evolution of crew resource management training in commercial aviation. *The International Journal of Aviation Psychology*, *9*(1), 19–32. https://www.taylorfrancis.com/books/9781315066264

Hemingway, M. W., O'Malley, C., & Silvestri, S. (2015). Safety Culture and Care: A Program to Prevent Surgical Errors. *AORN Journal*, *101*(4), 404–415. https://doi.org/10.1016/j.aorn.2015.01.002

Henkin, S., Chon, T. Y., Christopherson, M. L., Halvorsen, A. J., Worden, L. M., & Ratelle, J. T. (2016). Improving nurse–physician teamwork through interprofessional bedside rounding. *Journal of Multidisciplinary Healthcare*, *9*, 201–205. https://doi.org/10.2147/JMDH.S106644

Higgins, J. P. T., Altman, D. G., Gotzsche, P. C., Juni, P., Moher, D., Oxman, A. D., Savovic, J., Schulz, K. F., Weeks, L., & Sterne, J. A. C. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, *343*(oct18 2), d5928–d5928. https://doi.org/10.1136/bmj.d5928

Hiller, J. H., Fisher, G. A., & Kaess, W. (1969). A Computer Investigation of Verbal Characteristics of Effective Classroom Lecturing. *American Educational Research Journal*, *6*(4), 661–675. https://doi.org/10.3102/00028312006004661

Hirschman, A. O. (1970). *Exit, voice and Loyalty: Responses to Decline in Firms, Organisations, and States*. Harvard University Press.

Hodges, S. (2018). *The Effect of Relationship Strength on Safety Voicing* [University of Canterbury]. http://hdl.handle.net/10092/15072

Hofmann, D. A., Morgeson, F. P., & Gerras, S. J. (2003). Climate as a moderator of the relationship between leader-member exchange and content specific citizenship: Safety climate as an exemplar. *The Journal of Applied Psychology*, *88*(1), 170–178. https://doi.org/10.1037/0021-9010.88.1.170

Hofstede, G. (2002). Dimensions Do Not Exist: A Reply to Brendan McSweeney. *Human Relations*, *55*(11), 1355–1361. https://doi.org/10.1177/0018726702055511004

Hofstede, G. (2015). *Dimension data matrix*. https://geerthofstede.com/wp-content/uploads/2016/08/6-dimensions-for-website-2015-08-16.csv

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and Organizations: Software of the Mind, revised and expanded 3rd ed*. McGraw-Hill, New York, NY.

Hogarth, R. M., Portell, M., & Cuxart, A. (2007). What risks do people perceive in everyday life? A perspective gained from the experience sampling method (ESM). *Risk Analysis*, *27*(6), 1427–1439. https://doi.org/10.1111/j.1539-6924.2007.00978.x

Hosseinian, S. S., & Torghabeh, Z. J. (2012). Major Theories of Construction Accident Causation Models: a Literature Review. *International Journal of Advances in Engineering & Technology*, *4*(2), 2231–1963.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (2004). Leadership, culture, and organizations: The GLOBE study of 62 societies. In *Beverly Hills, CA: Sage Publications*.

Hovey, R. B., Dvorak, M. L., Burton, T., Worsham, S., Padilla, J., Hatlie, M. J., & Morck, A. C. (2011). Patient Safety: A Consumer's Perspective. *Qualitative Health Research*, *21*(5), 662–672. https://doi.org/10.1177/1049732311399779

Howard, M., Fleming, M. L., & Parker, E. (2013). Patients Do Not Always Complain When They Are Dissatisfied. *Journal of Patient Safety*, *9*(4), 224–231. https://doi.org/10.1097/PTS.0b013e3182913837

Hu, Y.-Y., Parker, S. H., Lipsitz, S. R., Arriaga, A. F., Peyre, S. E., Corso, K. A., Roth, E. M., Yule, S. J., &

Greenberg, C. C. (2016). Surgeons' Leadership Styles and Team Behavior in the Operating Room. *Journal of the American College of Surgeons*, *222*(1), 41–51. https://doi.org/10.1016/j.jamcollsurg.2015.09.013

Huang, X., Van de Vliert, E., & Van der Vegt, G. (2005). Breaking the Silence Culture: Stimulation of Participation and Employee Opinion Withholding Cross-nationally. *Management and Organization Review*, *1*(3), 459–482. https://doi.org/10.1111/j.1740-8784.2005.00023.x

Hughes, K. M., Benenson, R. S., Krichten, A. E., Clancy, K. D., Ryan, J. P., & Hammond, C. (2014). A Crew Resource Management Program Tailored to Trauma Resuscitation Improves Team Behavior and Communication. *Journal of the American College of Surgeons*, *219*(3), 545–551. https://doi.org/10.1016/j.jamcollsurg.2014.03.049

Islam, G., & Zyphur, M. J. (2005). Power, Voice, and Hierarchy: Exploring the Antecedents of Speaking Up in Groups. *Group Dynamics: Theory, Research, and Practice*, *9*(2), 93–103. https://doi.org/10.1037/1089-2699.9.2.93

Jackson, D., Peters, K., Andrew, S., Edenborough, M., Halcomb, E., Luck, L., Salamonson, Y., & Wilkes, L. (2010). Understanding whistleblowing: qualitative insights from nurse whistleblowers. *Journal of Advanced Nursing*, *66*(10), 2194–2201. https://doi.org/10.1111/j.1365-2648.2010.05365.x

Jacobsson, M., Hargestam, M., Hultin, M., & Brulin, C. (2012). Flexible knowledge repertoires: communication by leaders in trauma teams. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, *20*(1), 44. https://doi.org/10.1186/1757-7241-20-44

James, W. (1910). *Essays in radical empiricism*. Harvard University Press.

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*, *2*(4), 230–243. https://doi.org/10.1136/svn-2017-000101

Jiménez-Buedo, M., & Miller, L. M. (2010). Why a Trade-Off? the Relationship Between the External and Internal Validity of Experiments. *An International Journal for Theory*, *25*(3), 301–321. http://www.ehu.es/ojs/index.php/THEORIA/article/view/779/700

Johnson, H. L., & Kimsey, D. (2012). Patient Safety: Break the Silence. *AORN Journal*, *95*(5), 591–601. https://doi.org/http://dx.doi.org/10.1016/j.aorn.2012.03.002

Johnson, K. R., Heilig, L. F., Hester, E. J., Francis, S. O., Deakyne, S. J., & Dellavalle, R. P. (2006). Indoor tanning attitudes and practices of US dermatologists compared with other medical specialists. *Archives of Dermatology*, *142*(4), 465–470. https://doi.org/10.1001/archderm.142.4.465

Jonas, E., McGregor, I., Klackl, J., Agroskin, D., Fritsche, I., Holbrook, C., Nash, K., Proulx, T., & Quirin, M. (2014). Threat and Defense. From Anxiety to Approach. In *Advances in Experimental Social Psychology* (1st ed., Vol. 49). Elsevier Inc. https://doi.org/10.1016/B978-0-12-800052-6.00004-4

Jones, A., & Kelly, D. (2014). Deafening silence? Time to reconsider whether organisations are silent or deaf when things go wrong. *BMJ Quality & Safety*, *July*, 1–5. https://doi.org/10.1136/bmjqs-2013-002718

Jones, A., Lankshear, A., & Kelly, D. (2016). Giving voice to quality and safety matters at board level: A qualitative study of the experiences of executive nurses working in England and Wales. *International Journal of Nursing Studies*, *59*, 169–176. https://doi.org/10.1016/j.ijnurstu.2016.04.007

Jones, C., & Durbridge, M. (2016). Culture, silence and voice: The implications for patient safety in the operating theatre. *Journal of Perioperative Practice*, *26*(12), 281–284. https://doi.org/10.1177/175045891602601204

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: a failure to disagree. *The American Psychologist*, *64*(6), 515–526. https://doi.org/10.1037/a0016755

Kam, C. C.-S., & Bond, M. H. (2009). Emotional reactions of anger and shame to the norm violation characterizing episodes of interpersonal harm. *British Journal of Social Psychology*, *48*(2), 203–219. https://doi.org/10.1348/014466608X324367

Kanki, B., Anca, J., & Chidester, T. (2019). *Crew Resource Management* (3rd ed.). Elsevier Inc.

Kaposi, D. (2017). The resistance experiments: Morality, authority and obedience in Stanley Milgram's account. *Journal for the Theory of Social Behaviour*, *47*(4), 382–401. https://doi.org/10.1111/jtsb.12137

Kassing, J. W. (2002). Speaking up: Identifying Employees' Upward Dissent Strategies. *Management Communication Quarterly*, *16*(2), 187–209. https://doi.org/10.1177/089331802237234

Kawahara, H., & Morise, M. (2011). Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework. *Sadhana*, *36*(5), 713–727. https://link.springer.com/content/pdf/10.1007/s12046-011-0043-3.pdf

Kendrick, K. H. (2017). Using Conversation Analysis in the Lab. *Research on Language and Social Interaction*, *50*(1), 1–11. https://doi.org/10.1080/08351813.2017.1267911

Kenny, K., Fotaki, M., & Vandekerckhove, W. (2020). Whistleblower Subjectivities: Organization and Passionate Attachment. *Organization Studies*, *41*(3), 323–343. https://doi.org/10.1177/0170840618814558

Kines, P., Andersen, L. P. S. S., Spangenberg, S., Mikkelsen, K. L., Dyreborg, J., & Zohar, D. (2010). Improving construction site safety through leader-based verbal safety communication. *Journal of Safety Research*, *41*(5), 399–406. https://doi.org/http://dx.doi.org/10.1016/j.jsr.2010.06.005

King, H. B., Battles, J., Baker, D. P., Alonso, A., Salas, E., Webster, J., Toomey, L., & Salisbury, M. (2008). TeamSTEPPS™: team strategies and tools to enhance performance and patient safety. In K. Henriksen, J. Battles, M. Keyes, & M. Grady (Eds.), *Advances in Patient Safety: new directions and alternative approaches (Vol. 3: performance and tools)*. Agency for Healthcare Research and Quality (US). https://www.ncbi.nlm.nih.gov/books/NBK43686/

Kitayama, S., Park, H., Sevincer, A. T., Karasawa, M., & Uskul, A. K. (2009). A cultural task analysis of implicit independence: Comparing North America, Western Europe, and East Asia. *Journal of Personality and Social Psychology*, *97*(2), 236–255. https://doi.org/10.1037/a0015999

Klaminder, J., Hellström, G., Fahlman, J., Jonsson, M., Fick, J., Lagesson, A., Bergman, E., & Brodin, T. (2016). Drug-Induced Behavioral Changes: Using Laboratory Observations to Predict Field Observations. *Frontiers in Environmental Science*, *4*(December), 1–9. https://doi.org/10.3389/fenvs.2016.00081

Kohn, L. T. E., Corrigan, J. M. E., & Donaldson, M. S. E. (1999). Executive summary. To err is human: building a safer health system. In *Institute of Medicine Committee on Quality of Health Care in America*. https://doi.org/10.1017/S095026880100509X

Kolbe, M., Burtscher, M. J., & Manser, T. (2013). Co-ACT-A framework for observing coordination behaviour in acute care teams. *BMJ Quality and Safety*, *22*(7), 596–605. https://doi.org/10.1136/bmjqs-2012-001319

Kolbe, M., Burtscher, M. J., Wacker, J., Grande, B., Nohynkova, R., Manser, T., Spahn, D. R., & Grote, G. (2012). Speaking up is related to better team performance in simulated anesthesia inductions: An observational study. *Anesthesia and Analgesia*, *115*(5), 1099–1108. https://doi.org/10.1213/ANE.0b013e318269cd32

Kolbe, M., Grande, B., & Spahn, D. R. (2015). Briefing and debriefing during simulation-based training and beyond: Content, structure, attitude and setting. *Best Practice & Research Clinical Anaesthesiology*, *29*(1), 87–96. https://doi.org/http://dx.doi.org/10.1016/j.bpa.2015.01.002

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

Kolbe, M., Weiss, M., Grote, G., Knauth, A., Dambach, M., Spahn, D. R., & Grande, B. (2013). TeamGAINS: A tool for structured debriefings for simulation-based team trainings. *BMJ Quality and Safety*, *22*(7), 541–553. https://doi.org/10.1136/bmjqs-2012-000917

Konrath, S., Meier, B. P., & Bushman, B. J. (2018). Development and validation of the single item trait empathy scale (SITES). *Journal of Research in Personality*, *73*(November), 111–122.

https://doi.org/10.1016/j.jrp.2017.11.009

Krenz, H., Burtscher, M. J., Grande, B., & Kolbe, M. (2020). Nurses' voice: the role of hierarchy and leadership. *Leadership in Health Services*, *33*(1), 12–26. https://doi.org/10.1108/LHS-07-2019-0048

Krenz, H., Burtscher, M. J., & Kolbe, M. (2019). "Not only hard to make but also hard to take:" Team leaders' reactions to voice. *Gruppe. Interaktion. Organisation. Zeitschrift Fur Angewandte Organisationspsychologie*, *50*(1), 3–13. https://doi.org/10.1007/s11612-019-00448-2

Krueger, J. I., & Massey, A. L. (2009). A rational reconstruction of misbehavior. *Social Cognition*, *27*(5), 786–812. https://doi.org/10.1521/soco.2009.27.5.786

Kulig, A. W., & Blanchard, R. D. (2016). Use of Cognitive Simulation During Anesthesiology Resident Applicant Interviews to Assess Higher-Order Thinking. *Journal of Graduate Medical Education*, *8*(3), 417–421. https://doi.org/10.4300/JGME-D-15-00367.1

Kurzon, D. (2007). Towards a typology of silence. *Journal of Pragmatics*, *39*(10), 1673–1688. https://doi.org/10.1016/j.pragma.2007.07.003

Kurzon, D. (2011). On silence. *Journal of Pragmatics*, *43*(9), 2275–2277. https://doi.org/10.1016/j.pragma.2010.11.011

Kwon, B., Farndale, E., & Park, J. G. (2016). Employee voice and work engagement: Macro, meso, and micro-level drivers of convergence? *Human Resource Management Review*, *26*(4), 327–337. https://doi.org/10.1016/j.hrmr.2016.04.005

Landau, J. (2009). To speak or not to speak: Predictors of voice propensity. *Journal of Organizational Culture, Communications and Conflict*, *13*(1), 35–54. https://search.proquest.com/docview/216598748?accountid=9630

Lappi, O., Rinkkala, P., & Pekkanen, J. (2017). Systematic observation of an expert driver's gaze strategy-An on-road case study. *Frontiers in Psychology*, *8*(April), 1–15. https://doi.org/10.3389/fpsyg.2017.00620

Lars, S., & Wolf, O. T. (2009). The context counts: Congruent learning and testing environments prevent memory retrieval impairment following stress. *Cognitive, Affective and Behavioral Neuroscience*, *9*(3), 229–236. https://doi.org/10.3758/CABN.9.3.229

Latane, B., & Darley, J. M. (1968). Group inhibition of bystander intervention in emergencies. *Journal of Personality and Social Psychology*, *10*(3), 215–221. https://doi.org/10.1037/h0026570

Latané, B., & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?* Appleton-Century-Crofts.

Lee-Treweek, G., & Linkogle, S. (2000). *Danger in the field: Risk and ethics in social research*. Routledge.

Legg, C., & Hookway, C. (2019). Pragmatism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2019/entries/pragmatism/

Leonard, M., Graham, S., & Bonacum, D. (2004). The human factor: the critical importance of effective teamwork and communication in providing safe care. *Quality and Safety in Health Care*, *13*(SUPPL. 1), 85–90. https://doi.org/10.1136/qshc.2004.010033

LePine, J. A., & Van Dyne, L. (1998). Predicting voice behavior in work groups. *Journal of Applied Psychology*, *83*(6), 853–868. https://doi.org/10.1037/0021-9010.83.6.853

Leveson, N. G. (2002). *System Safety Engineering : Back To The Future*. Massachusetts Institute of Technology.

Leveson, N. G. (2011). *Engineering a safer world: Systems thinking applied to safety*. MIT press.

Liang, J., Farh, C. I. C., & Farh, J.-L. (2012). Psychological Antecedents of Promotive and Prohibitive Voice: A Two-Wave Examination. *Academy of Management Journal*, *55*(1), 71–92. https://doi.org/10.5465/amj.2010.0176

Liao, J. M., Thomas, E. J., & Bell, S. K. (2014). Speaking up about the dangers of the hidden curriculum. *Health Affairs*, *33*(1), 168–171. https://doi.org/10.1377/hlthaff.2013.1073

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*, *6*(7), 1–28. https://doi.org/10.1371/journal.pmed.1000100

Lin, S.-H., & Johnson, R. E. (2014). Promotive and Prohibitive Voice Behaviors: The Role of Self-Regulation. *Academy of Management Proceedings*, *2014*(1), 12879. https://doi.org/10.5465/AMBPP.2014.12879abstract

Lin, S.-H., & Johnson, R. E. (2015). A Suggestion to Improve a Day Keeps Your Depletion Away: Examining Promotive and Prohibitive Voice Behaviors Within a Regulatory Focus and Ego Depletion Framework. *Journal of Applied Psychology*, *100*(5), 1381–1397. https://doi.org/http://dx.doi.org/10.1037/apl0000018

Lindberg, E., Persson, E., Hörberg, U., & Ekebergh, M. (2013). Older patients' participation in team meetings—A phenomenological study from the nurses' perspective. *International Journal of Qualitative Studies on Health and Well-Being*, *8*(1), 1–9. https://doi.org/10.3402/qhw.v8i0.21908

Lindell, M. K. (2018). Communicating Imminent Risk. In *Contemporary Sociology: A Journal of Reviews* (Vol. 37, Issue 2, pp. 449–477). https://doi.org/10.1007/978-3-319-63254-4_22

Lindell, M. K., & Perry, R. W. (2012). The Protective Action Decision Model: Theoretical Modifications and Additional Evidence. *Risk Analysis*, *32*(4), 616–632. https://doi.org/10.1111/j.1539-6924.2011.01647.x

Linstone, H. A., & Turoff, M. (1975). *The Delphi method: Techniques and Applications*. Addison-Wesley. https://doi.org/10.1007/s00256-011-1145-z

Liu, P., & Ma, L. (2016). Food scandals, media exposure, and citizens' safety concerns: A multilevel analysis across Chinese cities. *Food Policy*, *63*, 102–111. https://doi.org/http://dx.doi.org/10.1016/j.foodpol.2016.07.005

Liu, Wei, Gerdtz, M., & Manias, E. (2016). Creating opportunities for interdisciplinary collaboration and patient-centred care: how nurses, doctors, pharmacists and patients use communication strategies when managing medications in an acute hospital setting. *Journal of Clinical Nursing*, *25*(19–20), 2943–2957. https://doi.org/10.1111/jocn.13360

Liu, Wu, Zhu, R., & Yang, Y. (2010). I warn you because I like you: Voice behavior, employee identifications, and transformational leadership. *Leadership Quarterly*, *21*(1), 189–202. https://doi.org/10.1016/j.leaqua.2009.10.014

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, *127*(2), 267–286. https://doi.org/10.1037/0033-2909.127.2.267

LSE. (2016a). *The Ethics Code*. 4. https://www.cranfield.ac.uk/~/media/files/corporate_documents/ethics-code.ashx?la=en

LSE. (2019). *Code of Research Conduct* (Vol. 16, Issue July). https://info.lse.ac.uk/staff/Services/Policies-and-procedures/Assets/Documents/codResCon.pdf

LSE. (2016b). *Ethics guidance: A companion to the Ethics Code*. 1–7. https://info.lse.ac.uk/staff/divisions/Governance-Legal-and-Policy/Assets/Documents/Ethics/Ethics-Code-Guidance-Sept-2016.pdf

Lupton, D. (1999). *Risk: key ideas*. Routledge.

Lyndon, A. (2008). Social and environmental conditions creating fluctuating agency for safety in two urban academic birth centers. *Journal of Obstetric, Gynecologic, and Neonatal Nursing*, *37*(1), 13–23. https://doi.org/10.1111/j.1552-6909.2007.00204.x

Ma, C. (2016). The role of approach and avoidance motivation in employee voice. *International Journal of Productivity and Performance Management*, *65*(6), 744–759. https://doi.org/10.1108/IJPPM-01-2016-0002

MacNab, B., Brislin, R., Worthley, R., Galperin, B. L., Jenner, S., Lituchy, T. R., MacLean, J., Aguilera, G. M., Ravlin, E., Tiessen, J. H., Bess, D., & Turcotte, M. F. (2007). Culture and ethics management: Whistle-blowing and internal reporting within a NAFTA country context. *International Journal of Cross Cultural Management*, *7*(1), 5–28. https://doi.org/10.1177/1470595807075167

Maher, R. C. (2018). Application Example 2: Cockpit Voice Recorders. In R. C. Maher (Ed.), *Principles of Forensic Audio Analysis* (pp. 137–142). Springer Nature Switzerland. https://doi.org/10.1007/978-3-319-99453-6_10

Malloy, D., Hadjistavropoulos, McCarthy, E., Evans, R., Zakus, D., Park, Lee, & Williams. (2009). Culture and organizational climate: nurses' insights into their relationship with physicians. *Nursing Ethics*, *16*(6), 719–733. https://doi.org/10.1177/096973300934263

Malvey, D., Fottler, M. D., & Sumner, J. (2013). The fear factor in healthcare: employee information sharing. *Journal of Healthcare Management*, *58*(3), 225–237. https://search.proquest.com/docview/1372122913?accountid=9630

Manapragada, A., & Bruk-Lee, V. (2016). Staying silent about safety issues: Conceptualizing and measuring safety silence motives. *Accident Analysis & Prevention*, *91*(November), 144–156. https://doi.org/10.1016/j.aap.2016.02.014

Manias, E. (2015). Communication relating to family members' involvement and understandings about patients' medication management in hospital. *Health Expectations*, *18*(5), 850–866. https://doi.org/10.1111/hex.12057

Manning, M. Lou. (2006). Patient Safety Improving Clinical Communication Through Structured Conversation. *Nursing*, *24*(5), 268–272.

Mannion, R., Blenkinsopp, J., Powell, M., McHale, J., Millar, R., Snowden, N., & Davies, H. (2018). Understanding the knowledge gaps in whistleblowing and speaking up in health care: narrative reviews of the research literature and formal inquiries, a legal analysis and stakeholder interviews. *Health Services and Delivery Research*, *6*(30), 1–190. https://doi.org/10.3310/hsdr06300

Mao, L. R. (1994). Beyond politeness theory: 'Face' revisited and renewed. *Journal of Pragmatics*, *21*(5), 451–486. https://doi.org/10.1016/0378-2166(94)90025-6

March, J. G., & Olsen, J. P. (1975). The Uncertainty of the Past: Organizational Learning Under Ambiguity. *European Journal of Political Research*, *3*(2), 147–171. https://doi.org/10.1111/j.1475-6765.1975.tb00521.x

Markiewiz, D. (2002). Speak up for safety: your views can make a difference. *Industrial Safety and Hygience*

News, *36*(7), 17–18. https://search.proquest.com/docview/196542598?accountid=9630

Marshall, K. G. (1996). Prevention. How much harm? How much benefit? 3. Physical, psychological and social harm. *Canadian Medical Association Journal*, *155*(2), 169–176. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1487962/

Mastrofski, S. D., Parks, R. B., & McCluskey, J. D. (2010). Systematic social observation in criminology. In *Handbook of quantitative criminology* (pp. 225–247). Springer. https://doi.org/10.1007/978-0-387-77650-7

Mastrofski, S. D., Parks, R. B., Reiss Jr., A. J., Worden, R. E., DeJong, C., Snipes, J. B., & Terrill, W. (1998). *Systematic Observation of Public Police: Applying Field Research Methods to Policy Issues*. https://www.ncjrs.gov/pdffiles/172859.pdf

Matthews, B., Andronaco, R., & Adams, A. (2014). Warning signs at beaches: Do they work? *Safety Science*, *62*, 312–318. https://doi.org/10.1016/j.ssci.2013.09.003

Maxfield, D. G., Lyndon, A., Kennedy, H. P., O'Keeffe, D. F., & Zlatnik, M. G. (2013). Confronting safety gaps across labor and delivery teams. *American Journal of Obstetrics and Gynecology*, *209*(5), 402–408. https://doi.org/10.1016/j.ajog.2013.07.013

McBee, M. (2010). Modeling Outcomes With Floor or Ceiling Effects: An Introduction to the Tobit Model. *Gifted Child Quarterly*, *54*(4), 314–320. https://doi.org/10.1177/0016986210379095

McLaughlin, N., Winograd, D., Chung, H. R., Van de Wiele, B., & Martin, N. A. (2014). Impact of the Time-Out Process on Safety Attitude in a Tertiary Neurosurgical Department. *World Neurosurgery*, *82*(5), 567–574. https://doi.org/http://dx.doi.org/10.1016/j.wneu.2013.07.074

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

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

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, *1*, 3111–3119. https://dwulff.github.io/_Naturallanguage/Literature/word2vec_b.pdf

Milgram, S. (1963). Behavioral Study of Obedience. *Journal of Abnormal Psychology*, *67*(4), 371–378. http://www.ncbi.nlm.nih.gov/pubmed/14049516

Milgram, S. (1974). *Obedience to Authority: An Experimental View.* Harper & Row.

Miller, A. G. (1986). *The obedience experiments: A case study of controversy in social science.* http://psycnet.apa.org/record/1986-97820-000

Miller, A. G., Collins, B. E., & Brief, D. E. (1995). Perspectives on Obedience to Authority: The Legacy of the Milgram Experiments. *Journal of Social Issues*, *51*(3), 1–19. https://doi.org/10.1111/j.1540-4560.1995.tb01331.x

Mitchell, G. (2012). Revisiting Truth or Triviality: The External Validity of Research in the Psychological Laboratory. *Perspectives on Psychological Science*, *7*(2), 109–117. https://doi.org/10.1177/1745691611432343

Moorhead, G., Ference, R., & Neck, C. P. (1991). Group Decision Fiascoes Continue: Space Shuttle Challenger and a Revised Groupthink Framework. *Human Relations*, *44*(6), 539–550. https://doi.org/10.1177/001872679104400601

Morison, T., & Macleod, C. (2014). When veiled silences speak: reflexivity, trouble and repair as methodological tools for interpreting the unspoken in discourse-based data. *Qualitative Research*, *14*(6), 694–711. https://doi.org/10.1177/1468794113488129

Morrison, E. W. (2011). Employee Voice Behavior: Integration and Directions for Future Research. *The 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

Morrison, E. W., See, K. E., & Pan, C. (2015). An Approach-Inhibition Model of Employee Silence: The Joint Effects of Personal Sense of Power and Target Openness. *Personnel Psychology*, *68*(3), 547–580. https://doi.org/10.1111/peps.12087

Morrison, E. W., Wheeler-Smith, S. L., & Kamdar, D. (2011). Speaking up in groups: a cross-level study of group voice climate and voice. *The Journal of Applied Psychology*, *96*(1), 183–191. https://doi.org/10.1037/a0020744

Morrow, K. J., Gustavson, A. M., & Jones, J. (2016). Speaking up behaviours (safety voices) of healthcare workers: A metasynthesis of qualitative research studies. *International Journal of Nursing Studies*, *64*(1), 42–51. https://doi.org/10.1016/j.ijnurstu.2016.09.014

Mowbray, P. K., Wilkinson, A., & Tse, H. H. M. (2015). An Integrative Review of Employee Voice: Identifying a Common Conceptualization and Research Agenda. *International Journal of Management Reviews*, *17*(3), 382–400. https://doi.org/10.1111/ijmr.12045

Mulhall, A. (2003). In the field: notes on observation in qualitative research. *Journal of Advanced Nursing*, *3*(41), 306–313. https://doi.org/10.1109/TBME.2009.2034734

Mullin, S. (2003). The anthrax attacks in New York City: the “Giuliani press conference model” and other communication strategies that helped. *Journal of Health Communication*, *8 Suppl 1*, 15–16. https://doi.org/10.1080/713851984

Mumford, C. (2015). *Withholding communication - a relational approach to silence in an inter-organisational project group in the UK* [Open University]. http://choicereviews.org/review/10.5860/CHOICE.51-2973

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

Murphy, E., & Dingwall, R. (2007b). Informed consent, anticipatory regulation and ethnographic practice. *Social Science & Medicine*, *65*(11), 2223–2234. https://doi.org/10.1016/j.socscimed.2007.08.008

Nabhan, M., Elraiyah, T., Brown, D. R., Dilling, J., Leblanc, A., Montori, V. M., Morgenthaler, T., Naessens, J., Prokop, L., Roger, V., Swensen, S., Thompson, R. L., & Murad, M. H. (2012). What is preventable harm in healthcare? A systematic review of definitions. *BMC Health Services Research*, *12*(1), 1–8. https://doi.org/10.1186/1472-6963-12-128

National Oil Spill Commission. (2011). *Deep water: The Gulf oil disaster and the future of offshore drilling*. National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling. https://digital.library.unt.edu/ark:/67531/metadc123527/m2/1/high_res_d/DEEPWATER_ReporttothePresident_FINAL.pdf

NATS. (2013). *Analysis of Infringements in NATS-Controlled Airspace*. http://flyontrack.co.uk/wp-content/uploads/2013/08/Analysis-1-201301.pdf

NATS. (2018a). *Advancing Aviation, keeping the skies safe*. https://www.nats.aero/wp-content/uploads/2016/08/NATS-Corp-Presentation-200418-V2.3.pptx

NATS. (2018b). *Facts, Stats & Reports: Company Performance 2015-16*. https://www.nats.aero/news/facts-stats-reports/

NATS. (2020). *The NATS Safety Strategy for 2030: The Future of Safety in ATM*. https://www.nats.aero/wp-content/uploads/2019/06/TheFutureOfSafetyInATM2030.pdf

Neal, A., Griffin, M. ., & Hart, P. . (2000). The impact of organizational climate on safety climate and individual behavior. *Safety Science*, *34*(1–3), 99–109. https://doi.org/10.1016/S0925-7535(00)00008-4

Neal, A., & Griffin, M. A. (2004). Safety climate and safety at work. In J. Barling & M. R. Frone (Eds.), *The* *psychology of workplace safety.* (pp. 15–34). American Psychological Association. https://doi.org/10.1037/10662-002

Near, J. P., & Miceli, M. P. (1985). Organizational dissidence: The case of whistle-blowing. *Journal of Business Ethics*, *4*(1), 1–16. https://doi.org/10.1007/BF00382668

Nembhard, I. M., & Edmondson, A. C. (2006). Making it safe: the effects of leader inclusiveness and professional status on psychological safety and improvement efforts in health care teams. *Journal of Organizational Behavior*, *27*(7), 941–966. https://doi.org/10.1002/job.413

Nembhard, I. M., Labao, I., & Savage, S. (2015). Breaking the silence. *Health Care Management Review*, *40*(3), 225–236. https://doi.org/10.1097/HMR.0000000000000028

Nembhard, I. M., Yuan, C. T., Shabanova, V., & Cleary, P. D. (2015). The relationship between voice climate and patients' experience of timely care in primary care clinics. *Health Care Management Review*, *40*(2), 104–115. https://doi.org/10.1097/HMR.0000000000000017

Nesse, R. M. (2005). Natural selection and the regulation of defenses. A signal detection analysis of the smoke detector principle. *Evolution and Human Behavior*, *26*(1), 88–105. https://doi.org/10.1016/j.evolhumbehav.2004.08.002

Nestel, D., Krogh, K., & Kolbe, M. (2017). Exploring realism in healthcare simulations. In *Healthcare Simulation Education* (pp. 23–28). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119061656.ch4

Nevile, M., & Walker, M. (2005). A context for error: using conversation analysis to represent and analyse recorded voice data. *Human Factors and Aerospace Safety*, *June*, 32. http://trid.trb.org/view.aspx?id=781567

Newman, A., Donohue, R., & Eva, N. (2017). Psychological safety: A systematic review of the literature. *Human Resource Management Review*, *27*(3), 521–535. https://doi.org/10.1016/j.hrmr.2017.01.001

Nichols, A. L., & Maner, J. K. (2008). The Good-Subject Effect: Investigating Participant Demand Characteristics. *The Journal of General Psychology*, *135*(2), 151–166. https://doi.org/10.3200/GENP.135.2.151-166

Noelle-Neumann, E. (1974). The Spiral of Silence a Theory of Public Opinion. *Journal of Communication*, *24*(2), 43–51. https://doi.org/10.1111/j.1460-2466.1974.tb00367.x

Noland, C. M., & Carmack, H. J. (2015). Narrativizing Nursing Students' Experiences With Medical Errors During Clinicals. *Qualitative Health Research*, *25*(10), 1423–1434. https://doi.org/10.1177/1049732314562892

Noort, M. C., Reader, T. W., & Gillespie, A. (2019a). Walking the Plank: An Experimental Paradigm to Investigate Safety Voice. *Frontiers in Psychology*, *10*(April), 1–18. https://doi.org/10.3389/fpsyg.2019.00668

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

Noort, M. C., Reader, T. W., & Gillespie, A. T. (2017). Prospero record: The Safety Voice Framework: A systematic review and comparison to employee voice. *PROSPERO CRD42017058264*, 1–11. https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=58264

Noort, M. C., Reader, T. W., Shorrock, S., & Kirwan, B. (2016). The relationship between national culture and safety culture: Implications for international safety culture assessments. *Journal of Occupational and Organizational Psychology*, *89*(3), 515–538. https://doi.org/10.1111/joop.12139

Novak, A. (2019). Improving safety through speaking up: An ethical and financial imperative. *Journal of Healthcare Risk Management*, *39*(1), 19–27. https://doi.org/10.1002/jhrm.21360

NTSB. (1978). *Aircraft Accident Report: United Airlines, Inc. McDonnell Douglas, Dc-8-61, N8082U*. http://libraryonline.erau.edu/online-full-text/ntsb/aircraft-accident-reports/AAR79-07.pdf

NTSB. (1983). *Aircraft Accident Report: Air Canada Flight 797*. http://libraryonline.erau.edu/online-full-text/ntsb/aircraft-accident-reports/AAR86-02.pdf

O'Connor, P., Byrne, D., O'Dea, A., McVeigh, T. P., & Kerin, M. J. (2013). Excuse me: Teaching interns to speak up. *Joint Commission Journal on Quality and Patient Safety*, *39*(9), 426–431. https://doi.org/10.1016/S1553-7250(13)39056-4

O'Connor, P., Campbell, J., Newon, J., Melton, J., Salas, E., & Wilson, K. A. (2008). Crew resource management training effectiveness: A meta-analysis and some critical needs. *International Journal of Aviation Psychology*, *18*(4), 353–368. https://doi.org/10.1080/10508410802347044

O'Connor, P., Jones, D. W., McCauley, M. E., & Buttrey, S. E. (2012). An evaluation of the effectiveness of the crew resource management programme in naval aviation. *International Journal of Human Factors and Ergonomics*, *1*(1), 21. https://doi.org/10.1504/ijhfe.2012.045272

O'Donovan, R., & McAuliffe, E. (2020). A systematic review exploring the content and outcomes of interventions to improve psychological safety, speaking up and voice behaviour. *BMC Health Services Research*, *20*(1), 101. https://doi.org/10.1186/s12913-020-4931-2

Okuyama, A., Wagner, C., & Bijnen, B. (2014). Speaking up for patient safety by hospital-based health care professionals: a literature review. *BMC Health Services Research*, *14*(1), 61. https://doi.org/10.1186/1472-6963-14-61

Omura, M., Maguire, J., Levett-Jones, T., & Stone, T. E. (2017). The effectiveness of assertiveness communication training programs for healthcare professionals and students: A systematic review. *International Journal of Nursing Studies*, *76*(August), 120–128. https://doi.org/10.1016/j.ijnurstu.2017.09.001

ONS. (2016). *Deaths from suicide, accidents and all other causes, 1979 to 2014 occurrences [006462]*. https://www.ons.gov.uk/file?uri=/peoplepopulationandcommunity/ birthsdeathsandmarriages/deaths/adhocs/006462deathsfromsuicideaccidentsandallothercauses1979to2014 occurrences/finalfile.xls

Orasanu, J., & Fischer, U. (1992). Team cognition in the cockpit: Linguistic control of shared problem solving. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society* (pp. 189–194). Lawrence Erlbauw Associates, Inc. https://books.google.co.uk/books?id=T7skQirc94IC&dq=Team+cognition+in+the+cockpit:+Linguistic+control+of+shared+problem+solving&lr=&source=gbs_navlinks_s

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Palmer, J. E. (2016). Recognizing the continuum of opportunities for third parties to prevent and respond to sexual assault and dating violence on a college campus. *Crime Prevention and Community Safety*, *18*(1), 1–18. https://doi.org/10.1057/cpcs.2015.18

Pander Maat, H., & Lentz, L. (2010). Improving the usability of patient information leaflets. *Patient Education and Counseling*, *80*(1), 113–119. https://doi.org/10.1016/j.pec.2009.09.030

Park, K., Park, J., Park, S. K., Kim, J., Kwon, S., Kwak, J., & Cha, M. (2013). Voice of the employees resonated in online bamboo forests. *In AAAI Workshop - Technical Report*, *WS-13-02*, 12–15. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/viewFile/6247/6302

Peadon, R. R., Hurley, J., & Hutchinson, M. (2020). Hierarchy and medical error : Speaking up when witnessing an error. *Safety Science*, *125*(January 2019), 104648. https://doi.org/10.1016/j.ssci.2020.104648

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015.* University of Texas at Austin. http://hdl.handle.net/2152/31333

Pennington, J., Socher, R., & Manning, C. D. (2014). *GloVe: Global Vectors for Word Representation*.

https://nlp.stanford.edu/projects/glove/

Pérez-Tejera, F., Valera, S., & Anguera, M. T. (2018). Using systematic observation and polar coordinates analysis to assess gender-based differences in park use in Barcelona. *Frontiers in Psychology*, *9*(NOV), 1–15. https://doi.org/10.3389/fpsyg.2018.02299

Perrow, C. (2011). *Normal accidents: Living with high risk technologies-Updated edition*. Princeton university press.

Phelps, M. E., & Reed, W. G. (2016). Improving hand hygiene compliance by changing safety culture in an academic medical center. *Canadian Journal of Infection Control*, *31*(4), 241–248. http://search.ebscohost.com/login.aspx?direct=true&db=rzh&AN=120397853&site=ehost-live

Pian-Smith, M. C. M., Simon, R., Minehart, R. D., Podraza, M., Rudolph, J., Walzer, T., & Raemer, D. (2009). Teaching residents the two-challenge rule: a simulation-based approach to improve education and patient safety. *Simulation in Healthcare : Journal of the Society for Simulation in Healthcare*, *4*(2), 84–91. https://doi.org/10.1097/SIH.0b013e31818cffd3

Plane Crash Info. (2019). *Last Words*. http://planecrashinfo.com/lastwords.htm

Podsakoff, P. M., & Organ, D. W. (1986). Self-Reports in Organizational Research: Problems and Prospects. *Journal of Management*, *12*(4), 531–544. https://doi.org/10.1177/014920638601200408

Purohit, H., Hampton, A., Shalin, V. L., Sheth, A. P., Flach, J., & Bhatt, S. (2013). What kind of #conversation is Twitter? Mining #psycholinguistic cues for emergency coordination. *Computers in Human Behavior*, *29*(6), 2438–2447. https://doi.org/10.1016/j.chb.2013.05.007

Raemer, D. B., Kolbe, M., Minehart, R. D., Rudolph, J. W., & Pian-Smith, M. C. M. (2016). Improving Anesthesiologists' Ability to Speak Up in the Operating Room. *Academic Medicine*, *91*(4), 530–539. https://doi.org/10.1097/ACM.0000000000001033

Rand, D. G., & Epstein, Z. G. (2014). Risking your life without a second thought: Intuitive decision-making and extreme altruism. *PLoS ONE*, *9*(10), 1–6. https://doi.org/10.1371/journal.pone.0109687

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, *5*, 1–12. https://doi.org/10.1038/ncomms4677

Reader, T. W., Flin, R., Mearns, K., & Cuthbertson, B. H. (2007). Interdisciplinary communication in the intensive care unit. *British Journal of Anaesthesia*, *98*(3), 347–352. https://doi.org/10.1093/bja/ael372

Reader, T. W., Gillespie, A., & Roberts, J. (2014). Patient complaints in healthcare systems: a systematic review and coding taxonomy. *BMJ Quality & Safety*, *23*(8), 678–689. https://doi.org/10.1136/bmjqs-2013-002437

Reader, T. W., Mearns, K., Lopes, C., & Kuha, J. (2016). Organizational support for the workforce and employee safety citizenship behaviors: A social exchange relationship. *Human Relations*. https://doi.org/10.1177/0018726716655863

Reader, T. W., Noort, M. C., Kirwan, B., Shorrock, S., & Kirwan, B. (2015). Safety sans frontières: An international safety culture model. *Risk Analysis*, *35*(5), 770–789. https://doi.org/10.1111/risa.12327

Reader, T. W., & O'Connor, P. (2014). The Deepwater Horizon explosion: non-technical skills, safety culture, and system complexity. *Journal of Risk Research*, *17*(3), 405–424. https://doi.org/10.1080/13669877.2013.815652

Reason, J. (1990). *Human error*. Cambridge university press.

Reason, J. (2000). Human error: models and management. *British Medical Journal*, *172*(June), 320–770. https://doi.org/10.1136/bmj.320.7237.768

Reime, M. H., Johnsgaard, T., Kvam, F. I., Aarflot, M., Breivik, M., Engeberg, J. M., & Brattebø, G. (2016). Simulated settings; powerful arenas for learning patient safety practices and facilitating transference to clinical practice. A mixed method study. *Nurse Education in Practice*, *21*, 75–82. https://doi.org/10.1016/j.nepr.2016.10.003

Reiss, A. J. (1971). Systematic Observation of Natural Social Phenomena. *Sociological Methodology*, *3*, 3. https://doi.org/10.2307/270816

Renn, O. (1992). Concepts of risk: a classification. In S. Krimsky (Ed.), *Social Theories of Risk* (pp. 53–79). Praeger. https://doi.org/10.18419/opus-7248

Renn, O. (1998). The role of risk perception for risk management. *Reliability Engineering and System Safety*, *59*(1), 49–62. https://doi.org/10.1016/S0951-8320(97)00119-1

Reynolds, B., & Seeger, M. W. (2005). Crisis and emergency risk communication as an integrative model. *Journal of Health Communication*, *10*(1), 43–55. https://doi.org/10.1080/10810730590904571

Rhee, J., Dedahanov, A., & Lee, D. (2014). Relationships among power distance, collectivism, punishment, and acquiescent, defensive, or prosocial silence. *Social Behavior and Personality*, *42*(5), 705–720. https://doi.org/10.2224/sbp.2014.42.5.705

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring Global Self-Esteem: Construct Validation of a Single-Item Measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, *27*(2), 151–161. https://doi.org/10.1177/0146167201272002

Rogers, W. P. ., Armstrong, N. A. ., Acheson, D. C. ., Covert, E. E. ., Feynman, R. P. ., Hotz, R. B. ., Kutyna, D. J. ., Ride, S. K. ., Rummel, R. W. ., & Sutter, J. F. (1986). *Report of the Presidential Commission on the Space Shuttle Challenger Accident*. https://spaceflight.nasa.gov/outreach/SignificantIncidents/assets/rogers_commission_report.pdf

Rusbult, C. E., Zembrodt, I. M., & Gunn, L. K. (1982). Exit, voice, loyalty, and neglect: Responses to dissatisfaction in romantic involvements. *Journal of Personality and Social Psychology*, *43*(6), 1230–1242. https://doi.org/10.1037/0022-3514.43.6.1230

Rydenfält, C., Johansson, G., Odenrick, P., Åkerman, K., & Larsson, P. A. (2015). Distributed leadership in the operating room: a naturalistic observation study. *Cognition, Technology and Work*, *17*(3), 451–460. https://doi.org/10.1007/s10111-014-0316-9

Sassen, C. (2005). *Linguistic Dimensions of Crisis Talk* (Vol. 136). John Benjamins Publishing Company. https://doi.org/10.1075/pbns.136

Schein, E. H., & Scheiner, P. (2016). Defining The Structure Of Culture. In E. H. Schein & P. Scheiner (Eds.), *Organizational Culture and Leadership* (5th ed., Issue 2016, pp. 1–76). Wiley.

Scheufle, D. A., & Moy, P. (2000). Twenty-Five Years of the Spiral of Silence: a Conceptual Review and Empirical Outlook. *International Journal of Public Opinion Research*, *12*(1), 3–28. https://doi.org/10.1093/ijpor/12.1.3

Schwappach, D. L. B., & Gehring, K. (2014a). Trade-offs between voice and silence: a qualitative exploration of oncology staff's decisions to speak up about safety concerns. *BMC Health Services Research*, *14*(1), 303. https://doi.org/10.1186/1472-6963-14-303

Schwappach, D. L. B., & Gehring, K. (2014b). 'Saying it without words': a qualitative study of oncology staff's experiences with speaking up about safety concerns. *BMJ Open*, *4*(5), e004740. https://doi.org/10.1136/bmjopen-2013-004740

Schwappach, D. L. B., & Gehring, K. (2014c). Silence That Can Be Dangerous: A Vignette Study to Assess Healthcare Professionals' Likelihood of Speaking up about Safety Concerns. *PLoS ONE*, *9*(8), e104720. https://doi.org/10.1371/journal.pone.0104720

Schwappach, D. L. B., & Gehring, K. (2015). Frequency of and predictors for withholding patient safety concerns among oncology staff: a survey study. *European Journal of Cancer Care*, *24*(3), 395–403. https://doi.org/10.1111/ecc.12255

Schwappach, D. L. B., & Richard, A. (2018). Speak up-related climate and its association with healthcare workers' speaking up and withholding voice behaviours: A cross-sectional survey in Switzerland. *BMJ Quality and Safety*, *27*(10), 836–843. https://doi.org/10.1136/bmjqs-2017-007388

Schwartz, M. S., Dunfee, T. W., & Kline, M. J. (2005). Tone at the top: An ethics code for directors? *Journal of Business Ethics*, *58*(1), 79–100. https://doi.org/10.1007/s10551-005-1390-y

Schwartz, S. H., & Gottlieb, A. (1976). Bystander reactions to a violent theft: Crime in Jerusalem. *Journal of Personality and Social Psychology*, *34*(6), 1188–1199. https://doi.org/10.1037/0022-3514.34.6.1188

Schwarz, N., Bless, H., Strack, F., Klumpp, G., & et al. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology*, *61*(2), 195–202. https://doi.org/10.1037//0022-3514.61.2.195

Searle, J. R. (1995). *The construction of social reality*. The Free Press.

Searle, J. R. (2008). Language and social ontology. *Theory and Society*, *37*(5), 443–459. https://doi.org/10.1007/s11186-008-9068-y

Seiden, S. C., Galvan, C., & Lamm, R. (2006). Enhancing patient safety. *Quality and Safety in Health Care*, *15*(4), 272–277. https://doi.org/10.1136/qshc.2006.018044

Sexton, J. B., & Helmreich, R. L. (2000). Analyzing Cockpit Communications: The Links Between Language, Performance, Error, and Workload. *Journal of Human Performance in Extreme Environments*, *5*(1). https://doi.org/10.7771/2327-2937.1007

Sexton, J. B., & Helmreich, R. L. (2003). Using language in the cockpit: Relationships with workload and performance. *Communication in High Risk Environments*, *12*, 57–74.

Sexton, J. B., Helmreich, R. L., Neilands, T. B., Rowan, K., Vella, K., Boyden, J., Roberts, P. R., & Thomas, E. J. (2006). The Safety Attitudes Questionnaire: psychometric properties, benchmarking data, and emerging research. *BMC Health Services Research*, *6*(1), 44. https://doi.org/10.1186/1472-6963-6-44

Shappell, S. a, & Wiegmann, D. a. (2000). The Human Factors Analysis and Classification System – HFACS. *USDOT/FAA/AM-00/7 Office of Aviation Medicine*, 19. https://doi.org/10.1177/1062860613491623

Sharpe, V. (1997). Why “do no harm”? *Theoretical Medicine*, *18*(1–2), 197–215. http://www.ncbi.nlm.nih.gov/pubmed/9129401

Sheeran, P. (2002). Intention — Behavior Relations : A Conceptual and Empirical Review. *European Review of Social Psychology*, *12*(1), 1–36. https://doi.org/10.1080/14792772143000003

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

Slovic, P. (1987). Perception of risk. *Science*, *236*(4799), 280–285. https://doi.org/10.1126/science.3563507

Slovic, P. (2016). Understanding perceived risk: 1978-2015. *Environment*, *58*(1), 25–29. https://doi.org/10.1080/00139157.2016.1112169

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2004). Risk as Analysis and Risk as Feelings: Some Thoughts about Affect, Reason, Risk, and Rationality. *Risk Analysis*, *24*(2), 311–322. https://doi.org/10.1111/j.0272-4332.2004.00433.x

Smith, B., & Searle, J. (2003). The Construction of Social Reality. An Exchange. *American Journal of Economics and Sociology*, *62*(1), 285–309. https://doi.org/10.1111/1536-7150.t01-1-00012

Smith, H. P. R. (1979). A simulator study of the interaction of pilot workload with errors, vigilance, and decisions. *Nasa Technical Memorandum*, *No. 78482*(January), 57. http://hdl.handle.net/2060/19790006598

Soeters, J. L., & Boer, P. C. (2000). Culture and Flight Safety in Military Aviation. *The International Journal of Aviation Psychology*, *10*(2), 111–133. https://doi.org/10.1207/S15327108IJAP1002

Somerville, M. (2006). Subjected bodies, or embodied subjects: Subjectivity and learning safety at work. *Work, Subjectivity and Learning*, *1994*, 37–52. http://link.springer.com/chapter/10.1007/1-4020-5360-6_3

Spruce, L. (2014). Back to Basics: Speak Up. *AORN Journal*, *99*(3), 407–415. https://doi.org/10.1016/j.aorn.2013.10.020

Stanton, N. A., Chambers, P. R. G., & Piggott, J. (2001). Situational awareness and safety. *Safety Science*, *39*(3), 189–204. https://doi.org/10.1016/S0925-7535(01)00010-8

Steup, M., & Neta, R. (2020). *Epistemology*. https://plato.stanford.edu/archives/sum2020/entries/epistemology

Stoffregen, T. A., Bardy, B. G., Smart, L. J., & Pagulayan, R. (2003). On the Nature and Evaluation of Fidelity in Virtual Environments. In L. J. Hettinger & M. Haas (Eds.), *Virtual and Adaptive Environments: Applications, Implications, and Human Performance Issues* (pp. 111–128). Lawrence Erlbauw Associates. https://doi.org/10.1201/9781410608888

Sundqvist, A. S., & Carlsson, A. A. (2014). Holding the patient's life in my hands: Swedish registered nurse anaesthetists' perspective of advocacy. *Scandinavian Journal of Caring Sciences*, *28*(2), 281–288. https://doi.org/10.1111/scs.12057

Swuste, P., van Gulijk, C., Groeneweg, J., Zwaard, W., Lemkowitz, S., & Guldenmund, F. (2020). From clapham junction to macondo, deepwater horizon: Risk and safety management in high-tech-high-hazard sectors: A review of English and Dutch literature: 1988–2010. *Safety Science*, *121*(November 2018), 249–282. https://doi.org/10.1016/j.ssci.2019.08.031

Tailstrike. (2019). *Cockpit Voice Recoder database*. http://tailstrike.com/database.htm

Tamuz, M., & Thomas, E. J. (2006). Classifying and interpreting threats to patient safety in hospitals: insights from aviation. *Journal of Organizational Behavior*, *27*(7), 919–940. https://doi.org/10.1002/job.419

Tarnow, E. (1999). Self-destructive obedience in the airplane cockpit and the concept of obedience optimization. In T. Blass (Ed.), *Obedience to authority* (pp. 125–138). Psychology Press.

Tarnow, E. (2000). Towards the Zero Accident Goal: Assisting the First Officer: Monitor and Challenge Captain Errors. *Journal of Aviation/Aerospace Education & Research*, *10*(1), 29–38. https://doi.org/10.15394/jaaer.2000.1269

Tost, L. P., Gino, F., & Larrick, R. P. (2012). Power, competitiveness, and advice taking: Why the powerful don't listen. *Organizational Behavior and Human Decision Processes*, *117*(1), 53–65. https://doi.org/10.1016/j.obhdp.2011.10.001

Tucker, S., Chmiel, N., Turner, N., Hershcovis, M. S., Stride, C. B., Sandy Hershcovis, M., & Stride, C. B. (2008). Perceived organizational support for safety and employee safety voice: The mediating role of coworker support for safety. *Journal of Occupational Health Psychology*, *13*(4), 319–330. https://doi.org/10.1037/1076-8998.13.4.319

Tucker, S., & Turner, N. (2011). Young worker safety behaviors: Development and validation of measures. *Accident Analysis & Prevention*, *43*(1), 165–175. https://doi.org/10.1016/j.aap.2010.08.006

Tucker, S., & Turner, N. (2014). Safety voice among young workers facing dangerous work: A policy-capturing approach. *Safety Science*, *62*, 530–537. https://doi.org/10.1016/j.ssci.2013.10.011

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

Tuncel, S., Lotlikar, H., Salem, S., & Daraiseh, N. (2006). Effectiveness of behaviour based safety interventions to reduce accidents and injuries in workplaces: Critical appraisal and meta-analysis. *Theoretical Issues in Ergonomics Science*, *7*(3), 191–209. https://doi.org/10.1080/14639220500090273

Turner, N., & Gray, G. C. (2009). Socially constructing safety. *Human Relations*, *62*(9), 1259–1266. https://doi.org/10.1177/0018726709339863

Turner, N., Tucker, S., & Kelloway, E. K. (2015). Prevalence and demographic differences in microaccidents and safety behaviors among young workers in Canada. *Journal of Safety Research*, *53*, 39–43. https://doi.org/10.1016/j.jsr.2015.03.004

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

Vadera, A. K., Pratt, M. G., & Mishra, P. (2013). Constructive Deviance in Organizations. *Journal of Management*, *39*(5), 1221–1276. https://doi.org/10.1177/0149206313475816

Valle, R. (2019). Toward a Psychology of Silence. *Humanistic Psychologist*, *47*(3), 219–261. https://doi.org/10.1037/hum0000120

van de Mortel, T. (2008). Faking it : social desirability response bias in self- report research report research. *Australian Journal of Advanced Nursing*, *25*(4), 40–48. http://www.ajan.com.au/ajan_25.4.html

van den Bos, K., & Lind, E. A. (2013). On Sense-Making Reactions and Public Inhibition of Benign Social Motives. In *Advances in Experimental Social Psychology* (1st ed., Vol. 48, pp. 1–58). Elsevier Inc. https://doi.org/10.1016/B978-0-12-407188-9.00001-6

van den Bos, K., Muller, P. a., & Damen, T. (2011). A Behavioral Disinhibition Hypothesis of Interventions in Moral Dilemmas. *Emotion Review*, *3*(3), 281–283. https://doi.org/10.1177/1754073911402369

van den Bos, K., Müller, P. A., & van Bussel, A. A. L. (2009). Helping to overcome intervention inertia in bystander's dilemmas: Behavioral disinhibition can improve the greater good. *Journal of Experimental Social Psychology*, *45*(4), 873–878. https://doi.org/10.1016/j.jesp.2009.03.014

van den Bos, K., van Lange, P. A. M., Lind, E. A., Venhoeven, L. a, Beudeker, D. a, Cramwinckel, F. M., Smulders, L., & van der Laan, J. (2011). On the benign qualities of behavioral disinhibition: because of the prosocial nature of people, behavioral disinhibition can weaken pleasure with getting more than you deserve. *Journal of Personality and Social Psychology*, *101*(4), 791–811. https://doi.org/10.1037/a0023556

van Dyne, L., Ang, S., Botero, I. C., & Dyne, L. Van. (2003). Conceptualizing employee silence and employee voice as multidimensional constructs. *Journal of Management Studies*, *40*(6), 1359–1392. https://doi.org/10.1111/1467-6486.00384

van Dyne, L., & LePine, J. A. (1998). Helping and voice extra-role behaviors: evidence of construct and predictive validity. *Academy of Management Journal*, *41*(1), 108–119. https://doi.org/10.2307/256902

van Heuven, V., & Boersma, P. (2001). Speak and unSpeak with PRAAT. *Glot International*, *5*(9–10), 341–347. https://www.fon.hum.uva.nl/paul/papers/speakUnspeakPraat_glot2001.pdf

van Schagen, I., & Sagberg, F. (2012). The potential benefits of naturalistic driving for road safety research: Theoretical and empirical considerations and challenges for the future. *Procedia-Social and Behavioral Sciences*, *48*, 692–701.

Vanderweele, T. J. (2009). On the distinction between interaction and effect modification. *Epidemiology*, *20*(6), 863–871. https://doi.org/10.1097/EDE.0b013e3181ba333c

Vaughan, D. (1996). *The Challenger launch decision: Risky technology, culture, and deviance at NASA*. University of Chicago press.

Vaughan, D. (1999). THE DARK SIDE OF ORGANIZATIONS: Mistake, Misconduct, and Disaster. *Annual Review of Sociology*, *25*(1), 271–305. https://doi.org/10.1146/annurev.soc.25.1.271

Vincent, C. (2010). Patient Safety: 2nd edition. In *Patient Safety: 2nd edition*. https://doi.org/10.1002/9781444323856

Viscusi, W. K., & Zeckhauser, R. J. (2015). Regulating Ambiguous Risks: The Less than Rational Regulation of Pharmaceuticals. *Journal of Legal Studies (0047-2530)*, *44*, S387–S422. https://doi.org/10.1086/684305

Vylomova, E., Rimell, L., Cohn, T., & Baldwin, T. (2016). Take and took, gaggle and goose, book and read: Evaluating the utility of vector differences for lexical relation learning. *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers*, *3*, 1671–1682. https://doi.org/10.18653/v1/p16-1158

Waller, M. J., & Kaplan, S. A. (2016). Systematic Behavioral Observation for Emergent Team Phenomena. *Organizational Research Methods*, *21*(2), 500–515. https://doi.org/10.1177/1094428116647785

Waring, J. J. (2005). Beyond blame: cultural barriers to medical incident reporting. *Social Science & Medicine*, *60*(9), 1927–1935. https://doi.org/10.1016/j.socscimed.2004.08.055

Weathington, B. L., Cunningham, C. J. L., & Pittenger, D. J. (2010). *Research methods for the behavioral and social sciences*. John Wiley & Sons.

Wei, X., Zhang, Z.-X., & Chen, X.-P. (2015). I Will Speak Up If My Voice Is Socially Desirable: A Moderated Mediating Process of Promotive Versus Prohibitive Voice. *Journal of Applied Psychology*, *100*(5), 1641–1652. https://doi.org/10.1037/a0039046

Weick, K. E. (1995). The Nature of Sensemaking. *Sensemaking in Organizations*, 1–62. https://doi.org/10.1017/CBO9781107415324.004

Weick, K. E. (2010). Reflections on enacted sensemaking in the bhopal disaster. *Journal of Management Studies*, *47*(3), 537–550. https://doi.org/10.1111/j.1467-6486.2010.00900.x

Weiss, M., Kolbe, M., Grote, G., Dambach, M., Marty, A., Spahn, D. R., & Grande, B. (2014). Agency and Communion Predict Speaking Up in Acute Care Teams. *Small Group Research*, *45*(3), 290–313. https://doi.org/10.1177/1046496414531495

Weiss, M., Kolbe, M., Grote, G., Spahn, D. R., & Grande, B. (2017). Why didn't you say something? Effects of after-event reviews on voice behaviour and hierarchy beliefs in multi-professional action teams. *European Journal of Work and Organizational Psychology*, *26*(1), 66–80. https://doi.org/10.1080/1359432X.2016.1208652

Weiss, M., Kolbe, M., Grote, G., Spahn, D. R., & Grande, B. (2018). We can do it! Inclusive leader language promotes voice behavior in multi-professional teams. *The Leadership Quarterly*, *29*(3), 389–402. https://doi.org/10.1016/j.leaqua.2017.09.002

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

Whicher, D. M., Kass, N. E., Audera-Lopez, C., Butt, M., Jauregui, I. L., Harris, K., Knoche, J., & Saxena, A. (2015). Ethical Issues in Patient Safety Research: A Systematic Review of the Literature. *Journal of Patient Safety*, *11*(3), 174–184. https://doi.org/10.1097/PTS.0000000000000064

White, D. (1995). Application of systems thinking to risk management: A review of the literature. *Management Decision*, *33*(10), 35–45. https://doi.org/10.1108/EUM0000000003918

Wiegmann, D. A., & Shappell, S. A. (2016). A Human Error Approach to Aviation Accident Analysis. In *A Human Error Approach to Aviation Accident Analysis*. Routledge.

Wilkinson, A., Barry, M., & Morrison, E. (2020). Toward an integration of research on employee voice. *Human Resource Management Review*, *30*(1), 100677. https://doi.org/10.1016/j.hrmr.2018.12.001

Wilkinson, A., Donaghey, J., Dundon, T., & Freeman, R. B. (2014). *Handbook of Research on Employee Voice: Elgar original reference*. Edward Elgar Publishing.

Wilson, R. (1979). Analyzing the daily risks of life. *Technology Review*, *81*(4), 41–46. http://users.physics.harvard.edu/~wilson/publications/pp212.pdf

Wongpakaran, N., Wongpakaran, T., Wedding, D., & Gwet, K. L. (2013). A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Medical Research Methodology*, *13*(1), 61–68. https://doi.org/10.1186/1471-2288-13-61

Zanko, M., & Dawson, P. (2012). Occupational Health and Safety Management in Organizations: A Review. *International Journal of Management Reviews*, *14*(3), 328–344. https://doi.org/10.1111/j.1468-2370.2011.00319.x

Zimbardo, P. G. (2011). Lucifer Effect. In *The Encyclopedia of Peace Psychology*. Blackwell Publishing Ltd. https://doi.org/10.1002/9780470672532.wbepp149

Zohar, D. (1980). Safety climate in industrial organizations: Theoretical and applied implications. *Journal of Applied Psychology*, *65*(1), 96–102. https://doi.org/10.1037/0021-9010.65.1.96

Zohar, D. (2010). Thirty years of safety climate research: Reflections and future directions. *Accident Analysis & Prevention*, *42*, 1517–1522. https://doi.org/10.1016/j.aap.2009.12.019

APPENDICES

The following appendices are provided as supplemental material to this thesis:

Appendix A. The systematic review protocol.

Appendix B. Ethics approval.

Appendix C. The Walking the Plank manual.

Appendix D. The Walking the Plank study materials.

Appendix E. The Jupyter notebook for Article 3.

Appendix A: Systematic review protocol

Search strategy

Research Questions	1) Which definitions, conceptualisations and theoretical backgrounds characterise safety voice? 2) What is the ecological nature of safety voice in terms of inhibiting (promoting) relationships between safety voice and antecedents, pragmatics, and outcomes across levels of analysis? 3) Which methodologies have been used for researching safety voice, how suitable are they, and what methodological challenges remain?
Researchers	Doctoral student (lead): Mark C. Noort Supervisor: Tom W. Reader Co-supervisor: Alex T. Gillespie
Study goals	Primary: Answer the research question listed above. Secondary: Evaluate the quality of primary studies on safety voice Summarise results from primary studies on safety voice Calculate an overall effect of predictors of safety voice (if a meta-analysis can be performed) Identify and understand any heterogeneity in the results Identify new research questions Identify the extent of the gap regarding the disinhibition of safety voice
Search steps	Identification: Search electronic databases for publications Screening: Select potential papers based on inclusion/exclusion criteria for title and abstracts. Eligibility: Access full text articles to assess eligibility. Request full-text copies of inaccessible papers from the LSE libraries, or authors. Hand-search: Search for additional articles using the reference list of key articles and published literature reviews.
Search engines	PubMed EBSCOhost
Search databases	Anthropology plus Business source complete CINAHL Plus with full text Communication and mass media complete Criminal justice abstracts with full text History of science, technology, medicine International political science abstracts MEDLINE Peace research abstracts PsycArticles PsycINFO SocINDEX with full text
Definitions	A discreet speech act to prevent physical harm, where the receiver might be willing and able to take action.
Search terms	Abstracts and titles need to have the combination of safety and voice. Given it involves safety in safety-critical team it is opted to not include synonyms of safety (e.g., harm, injury, etc.). A pilot search revealed that included papers included a variation of voice, speaking or silence. Safety voice

	Safe AND voic* *Safe* AND speak* *Safe* AND silenc*
Exclusion terms	Technology VoIP “Language skill?” Technology-mediated Computer PC Pitch Intonation Anatomy Hertz Vocal molecular Children “Mental health” HIV
Search timeframe	Articles published anytime.
Timing of searches	At the beginning of the literature review (March 2017) Prior to data analysis, to ensure up-to-date results (estimated: Mid-April 2017)
Screening and selection	Titles and abstracts of results retrieved through the electronic and hand searches will be screened for meeting the inclusion criteria by the lead author. Reliability will be ensured through independent coding of a subset (i.e., 15%) of the titles and abstracts by a second researcher. Disagreements will be resolved in discussion with all authors. If any amendments are required to the search strategy these will be implemented through repeating the screening process, until a satisfactory result is achieved.
Risk of bias (quality) assessment	PRISMA method (Liberati, et al., 2009)

Inclusion criteria

Types of papers	Include	Peer-reviewed articles
	Exclude	All other, i.e., Duplicates Book reviews Periodicals Editorials Literature reviews Dissertations On-going, unpublished trials Errata (unless they refer to extracted information) Conference proceedings (e.g., keynotes, panel discussions)
Language	Include	English
	Exclude	All other languages
Types of subject	Include	Studied whether people raise a safety concern to another person Behavioural
	Exclude	Voice technology (e.g., Voice over Internet Protocol: VoIP) Language skills Physiological voice Technology-mediated communication HIV/AIDS Medical diseases Primarily about ethics Primarily about law Intimate partner violence Studies on drugs or therapies, not focussing on raising concerns
Types of study	Include	Empirical, primary study Quantitative Qualitative *(Note: the number of Qual/Quant will be coded, to allow for a go/no-go decision on the meta-analysis. Qualitative papers will be included if there are insufficient papers and vice versa).*
	Exclude	No original research (data) Conceptual papers without data Authors advocate for a group People speak up when prompted (e.g., focus group) Call to speak up about an issue
Types of participants	Include	Interpersonal (i.e., OB) From/to all staff roles From/to all hierarchical positions
	Exclude	Non-adult participants Mental health patients Not unions or organisations advocating on an issue
Types of predictors	Include	All to be included.
	Exclude	None. No predictors discussed.
Types of outcome measures	Include	Outcomes of safety voice are discussed.
	Exclude	*No predictors of safety voice discussed. (?)*

Data extraction

Using a set Excel extraction form, the following information was extracted from included full-text articles. To ensure reliability, a subset (15%) of the included full-text articles was independently coded by a research assistant.	
Review information	Search ID Reviewer name
Article information	Authors Title Journal Year
Participants	Country (of study) Sample details (e.g., size, gender, age, demographics, etc.) Sample before and after exclusion criteria applied (e.g., outliers) Industry and organisation(s) Unit of analysis (i.e., individual, dyads, teams, organisations) Other Unique / part of longitudinal data set
Methods	Study design (e.g., Quant/Qual, experimental, interviews, etc.) Details of intervention/manipulation Operationalisation of safety voice: details and type (e.g., self/other-reported, objective)
Quality indicators	Cochrane quality indicators Comparability of study groups (+ inclusion of control group) Randomly selected/Convenience-based? Was randomisation performed blindly? Representativeness of sample Exclusions made (participants, settings, etc.) Treatment of missing data Ethics: approval and informed consent obtained
Theoretical Background	Definition of safety voice (literal) Theoretical paradigm used
Suggested predictors of safety voice	If possible, specify by: Individual Team Organisation Other
Study results	Effect sizes (i.e., Cohen's d, odds ratio, hazard ratio) Correlations b/B-values standard errors
Outcomes	Details on safety voice outcomes, if any.
Proposed intervention	Details on proposed interventions, if any.

Search histories

Search v3 (final)

Search	Terms (example from PubMED)	PubMED		EBSCO	
1 SAFETY VOICE	(*safe*[Title/Abstract]) AND silenc*[Title/Abstract]) OR (*safe*[Title/Abstract]) AND speak*[Title/Abstract]) OR (*safe*[Title/Abstract] AND voic*[Title/Abstract])	4263	15 March 2017 11:11	6,867	15 March 2017 11:46
2 Exclusions	Technolog*[Title/Abstract] OR VoIP[Title/Abstract] OR "Language skill?" [Title/Abstract] OR Technology-mediated[Title/Abstract] OR Computer?[Title/Abstract] OR PC[Title/Abstract] OR Pitch[Title/Abstract] OR Intonation[Title/Abstract] OR Anatomy[Title/Abstract] OR vocal[Title/Abstract] OR Child*[Title/Abstract] OR "Mental health"[Title/Abstract] OR Contraception[Title/Abstract] OR HIV[Title/Abstract] OR molecular[Title/Abstract] OR therapy[Title/Abstract] OR airway[Title/Abstract] OR syndrome[Title/Abstract] OR "-speaking"[Title/Abstract]	5272121	15 March 2017 11:12	4,108,922	15 March 2017 11:48
3	#1 NOT #2	1607	15 March 2017 11:12	3,788	15 March 2017 11:49
4 Filters	Limit to English (peer-reviewed: EBSCO) Note: EBSCO already removes exact duplicates.	1539	15 March 2017 11:13	3742 (1492)	15 March 2017 12:05

Search v2 (refined)

Search	Terms (example from PubMED)	PubMED		EBSCO	
1 SAFETY and Voice	(*safe*[Title/Abstract]) AND silenc*[Title/Abstract]) OR (*safe*[Title/Abstract]) AND speak*[Title/Abstract]) OR (*safe*[Title/Abstract] AND voic*[Title/Abstract])	4265	22 Feb. 17	6916	22 Feb. 17
2 ORG	team*[Title/Abstract] OR organi?ation*[Title/Abstract] OR leader*[Title/Abstract] OR manag*[Title/Abstract] OR supervis*[Title/Abstract] OR boss*[Title/Abstract] OR colleague*[Title/Abstract] OR partner*[Title/Abstract]	1363639	22 Feb. 17	7353626	22 Feb. 17
3	#1 AND #2	1051	22 Feb. 17	2823	22 Feb. 17
4 Exclusions	Technolog*[Title/Abstract] OR VoIP[Title/Abstract] OR "Language skill?"[Title/Abstract] OR Technology-mediated[Title/Abstract] OR Computer?[Title/Abstract] OR PC[Title/Abstract] OR Pitch[Title/Abstract] OR Intonation[Title/Abstract] OR Anatomy[Title/Abstract] OR vocal[Title/Abstract] OR Child*[Title/Abstract] OR "Mental health"[Title/Abstract] OR Contraception[Title/Abstract] OR HIV[Title/Abstract] OR molecular[Title/Abstract] OR therapy[Title/Abstract] OR airway[Title/Abstract] OR syndrome[Title/Abstract] OR "-speaking"[Title/Abstract]	5257446	22 Feb. 17	4096391	22 Feb. 17
5	#3 NOT #4	361	22 Feb. 17	1632	22 Feb. 17
6	Limit to English, Dutch, (peer-reviewed: EBSCO) Note: EBSCO already removes exact duplicates.	349	22 Feb. 17	759 (649)	22 Feb. 17

Search v1 (first formal search)

Search	Terms (example from PubMED)	PubMED		Web of Science		EBSCO host	
1 SAFETY	*safe*[Title/Abstract] OR *harm*[Title/Abstract] OR *hurt*[Title/Abstract] OR *injure*[Title/Abstract] OR *damage?[Title/Abstract] OR danger*[Title/Abstract] OR *secure[Title/Abstract] OR *risk?[Title/Abstract] OR *vulnerable[Title/Abstract] OR benign[Title/Abstract] OR concern[Title/Abstract]	<u>2987866</u>	20 Feb. 17	<u>12,558,674</u>	21 Feb. 17	1,887,404	21 Feb. 17
2 VOICE	Voice[Title/Abstract] OR Voice?[Title/Abstract] OR Voicing[Title/Abstract] OR Raise[Title/Abstract] OR Raise?[Title/Abstract] OR Raising[Title/Abstract] OR “speak up” [Title/Abstract] OR “speaks up” [Title/Abstract] OR “speaking up” [Title/Abstract] OR “spoke up” [Title/Abstract] OR “speak out” [Title/Abstract] OR “speaks out” [Title/Abstract] OR “speaking out” [Title/Abstract] OR “spoke out” [Title/Abstract] OR advocat*[Title/Abstract] OR whistle[Title/Abstract] OR Silen*[Title/Abstract] OR Quiet[Title/Abstract] OR Conceal*[Title/Abstract]	<u>282686</u>	20 Feb. 17	**<u>653,083</u>**	21 Feb. 17	786,750	21 Feb. 17
3 ORG	team*[Title/Abstract] OR organi?ation*[Title/Abstract] OR leader*[Title/Abstract] OR manag*[Title/Abstract] OR supervis*[Title/Abstract] OR boss*[Title/Abstract] OR colleague*[Title/Abstract] OR partner*[Title/Abstract]	<u>1363018</u>	20 Feb. 17	**<u>2,359,953</u>**	21 Feb. 17	4,579,437	21 Feb. 17
4	**#1 AND #2 AND #3**	<u>9562</u>	20 Feb. 17	**<u>12,329</u>**	21 Feb. 17	27,775	21 Feb. 17

5 Exclusions	Technolog*[Title/Abstract] OR VoIP[Title/Abstract] OR "Language skill?" [Title/Abstract] OR Technology-mediated[Title/Abstract] OR Computer?[Title/Abstract] OR PC[Title/Abstract] OR Pitch[Title/Abstract] OR Intonation[Title/Abstract] OR Anatomy[Title/Abstract] OR vocal[Title/Abstract] OR Child*[Title/Abstract] OR "Mental health"[Title/Abstract] OR Contraception[Title/Abstract] OR HIV[Title/Abstract] OR molecular[Title/Abstract] OR therapy[Title/Abstract] OR airway[Title/Abstract] OR syndrome[Title/Abstract] OR "-speaking"[Title/Abstract]	5374582	20 Feb. 17	**26,989,860**	21 Feb. 17	3,563,439	21 Feb. 17
6	**#4 NOT #5**	3533	20 Feb. 17	**6,206**	21 Feb. 17	20,618	21 Feb. 17
7	Limit to English, Dutch, (peer-reviewed: EBSCO)	3367	20 Feb. 17	**5,725**	21 Feb. 17	11,563	21 Feb. 17

Appendix B: Ethics approval

Departmental ethics application (pilot studies), November 2016.

Ethics Application
Department of Psychology and Behavioural Science

Title of project: Pilot study to test speaking up scenarios.

Name of Researcher(s): Mark C. Noort

Email Address: m.c.noort@lse.ac.uk

Name of Supervisor (for MSc/PhD projects): Tom Reader

Date: 4 November

		Yes	No	N/A
1	Will the proposed research entail any risk to the researcher(s)? (eg., entail travel to unstable regions, exposure to environmental risks, collection of sensitive data, or lone working in an unfamiliar context)		x	

If you ticked **Yes** to Q1, you should complete a **risk assessment form**

		Yes	No	N/A
2	Will you describe the main experimental procedures to participants in advance, so that they are informed about what to expect?		x	
3	Will you tell participants that their participation is voluntary?	X		
4	Will you obtain written consent for participation?	X		
5	If the research is observational, will you ask participants for their consent to being observed?	X		
6	Will you tell participants that they may withdraw at any time and for any reason?	X		
7	With questionnaires, will you give participants the option of omitting any questions they do not want to answer?	X		
8	Will you tell participants that their data will be treated with full confidentiality and that, if published, it will not be identifiable as theirs?	X		
9	Will you debrief participants at the end of their participation (i.e. given them a brief explanation of the study)?	x		

If you ticked **No** to any of Q2-9, you should **tick box B** overleaf.

			Yes	No	N/A
10	Will your project involve deliberately misleading participants in any way?		x		
11	Is there any realistic risk of you or any participants experiencing either physical or psychological distress or discomfort? If **Yes,** give details on a separate sheet and state what you will tell them to do if they should experience any problems (e.g., who they can contact for help).			x	
12	Does your project involve work with animals?			X	
13	Do participants fall into any of the following special groups? Note that you may also need to obtain satisfactory CRB clearance(or equivalent for overseas students).	Schoolchildren (under age 18)		X	
		People with learning or communication difficulties		X	
		Parents		X	
		People in custody		X	
		People engaged in illegal activities (e.g. drug taking)		X	

If you have ticked **Yes** to any of Q10-13 you should **tick box B** overleaf.

There is an obligation on the lead researcher or supervisor to bring to the attention of the Departmental Ethics Committee any issues with ethical implications not clearly covered by the above checklist.

PLEASE TICK **EITHER** BOX A OR BOX B BELOW AND **PROVIDE THE DETAILS REQUIRED** IN SUPPORT OF YOUR APPLICATION. THEN SIGN THE FORM.

	Tick box
A. I consider that this project has **no** significant ethical implications to be brought before the Departmental Ethics Committee	
Give a brief description of participants and procedure (methods, tests used etc.) in up to 150 words. *If you have ticked box A, then sign and submit this form (and any attachments) to the ISP Ethics Committee.*	

	Tick box
B. I consider that this project **may** have ethical implications that should be brought before the Departmental committee, and/or it will be carried out with children or other vulnerable populations	X

Please provide all the further information listed below on a separate attachment.

1. Title of project
2. Purpose of project and its academic rationale
3. Brief description of methods and measurements
4. Participants: recruitment methods, number, age, gender, exclusion/inclusion criteria
5. Consent, participant information, debriefing (***attach information, consent, & debrief sheets**)
6. A clear concise statement of ethical issues raised by the project and how you intend to deal with them.
7. Estimated start date and duration of the project.

If any of the above information is missing, your application will be returned to you.
If you have ticked box B, then sign and submit this form along with a separate document providing the above information (and any attachments) to the ISP Ethics Committee.

I am familiar with the BPS Guidelines for ethical practices in psychological research and I have discussed them with other researchers involved in the research (e.g., supervisor or co-researcher).

Student signature	Mark C Noort	Print Name	Mark C Noort	Date 23 11 2016
Supervisor signature		Print Name		Date

Statement of Ethical Approval: To be Completed by the Chair of the Ethics Committee

This project has been considered using agreed procedures and is now approved.

#000540: Ethics approval (Walking the Plank), March 2017

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Houghton Street
London WC2A 2AE
United Kingdom

tel: +44 (0)20 7106 1202
email: rescon@lse.ac.uk
www.lse.ac.uk

Research Division

Mark Christiaan Noort
Department of Psychological and Behavioural Science
m.c.noort@lse.ac.uk

17th March 2017

Dear Mark

Re: Experimental investigation of conditions that inhibit speaking up about safety [ref. 000540]

I refer to the above research proposal which you recently submitted for review by the Research Ethics Committee.

Having considered your ethics review application and supporting documents, I am satisfied that you have properly addressed the ethical issues raised by your proposed research. I am thus able in my capacity as Chair of the Committee to approve the application.

Please note that any significant changes to the research design must be reported to the Research Ethics Committee. Amendments to the research design that may affect participants and/or that may have ethical implications must be reviewed and approved by the Research Ethics Committee before commencement (or recommencement) of the project. The Research Ethics Committee may periodically conduct a selective audit of current research projects.

I would like to take this opportunity to wish you well with your research project.
If you have any further queries, please feel free to contact Lyn Grove, Research Division.

Yours sincerely,

Professor John Worrall
Chair of the Research Ethics Committee

cc. Lyn Grove, Research Division

#000540: Ethics approval (Amendment), September 2017

LSE THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Houghton Street
London WC2A 2AE
United Kingdom

tel: +44 (0)20 7106 1202
email: rescon@lse.ac.uk
www.lse.ac.uk

Research Division

Mark Christiaan Noort
Department of Psychological and Behavioural Science
m.c.noort@lse.ac.uk

29th September 2017

Dear Mark

Re: Amendment to study: Experimental investigation of conditions that inhibit speaking up about safety [ref. 000540]

I refer to the above research proposal which was approved by the Research Ethics Committee in March this year, and for which you have now submitted a proposed amendment (as detailed in your email of 27 September). I believe that the nature of the deception is not changed by your proposed change in protocol, and I am satisfied that you have the necessary safeguards in place to prevent any harm to participants. I am thus able in my capacity as Chair of the Committee to approve the application.

Please note that any significant changes to the research design must be reported to the Research Ethics Committee. Amendments to the research design that may affect participants and/or that may have ethical implications must be reviewed and approved by the Research Ethics Committee before commencement (or recommencement) of the project. The Research Ethics Committee may periodically conduct a selective audit of current research projects.

I would like to take this opportunity to wish you well with your research project.
If you have any further queries, please feel free to contact Lyn Grove, Research Division.

Yours sincerely,

JWorrall

Professor John Worrall
Chair of the Research Ethics Committee
cc. Lyn Grove, Research Division

#1051: Ethics approval (Cockpit Voice Recorder study), January 2020

Houghton Street
London WC2A 2AE
United Kingdom

tel: +44 (0)20 7852 3629
email: research.ethics@lse.ac.uk

www.lse.ac.uk

Research Ethics Committee

Mark Noort
Department of Psychological and Behavioural Science
m.c.noort@lse.ac.uk

27th January 2020

Dear Mark

Re: 'Analysis of online Cockpit Voice Recorder conversations of aviation crashes' [REC ref. 1051]

I am writing with reference to the above research proposal. The Research Ethics Committee, having considered the documentation sent, is satisfied that the ethical issues raised by the proposed research have been properly taken into account and that adequate safeguards have been put in place. I am accordingly able on behalf of the Committee to confirm our approval of the application.

Please note that any significant changes to the research design must be reported to the Research Ethics Committee. Amendments to the research design that may affect participants and/or that may have ethical implications must be reviewed and approved by the Research Ethics Committee before commencement (or recommencement) of the project. The Research Ethics Committee may periodically conduct a selective audit of current research projects.

I would like to take this opportunity to wish you well with your research project.
If you have any further queries, please feel free to contact Lyn Grove, Research Division.

Yours sincerely,

Dr Ilka Gleibs
Deputy Chair, Research Ethics Committee
cc. Dr Lyn Grove, Research Division

Appendix C: Walking the Plank manual

Introduction[36]

'Safety voice' is the act of speaking-up about safety issues. It is defined as "explicit communication that is 1) discretionary, 2) aimed at improving a perceived unsafe situation, and 3) addressed to others of equal or senior status" (Noort et al., 2019b) and through speaking up about safety, hazards can be identified and mitigated.

This manual outlines the protocol for the investigation of safety voice through using the Walking the Plank paradigm. Five stages are required to enable the direct observation of safety voice behaviours, and these involve i) participant welcome and informed consent, ii) creativity task, iii) demonstration and evaluation of creative ideas, iv) wrap-up questionnaire, and v) debrief. To enable interpretation and amendments to the protocol, and successful execution of the protocol, we have provided detailed information for each stage of the Walking the Plank paradigm, a checklist for materials and key decisions, appendices to illustrate study materials, illustrative pictures.

The Walking the Plank paradigm makes accessible safety voice behaviours, and you are encouraged to tailor procedures for the purpose of your research. Researching safety voice is

[36] This appendix has been reformatted for a consistent lay-out within the thesis. The original manual is published as supplementary material to Article 2 and can bound online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00668/full#supplementary-material

fascinating, and we trust that future research will find this paradigm useful for identifying novel concepts for speaking-up about safety, and new ways for people to create safety.

Mark C. Noort

And on behalf of:

Dr Tom W. Reader,

Dr Alex Gillespie.

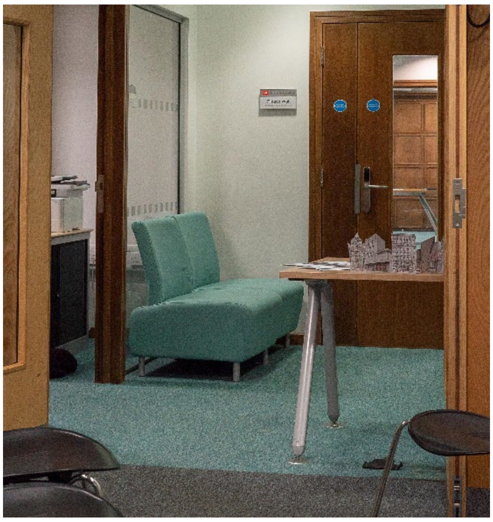
Walking the Plank: an overview

The Walking the plank paradigm consists of 3 core stages: i) a creativity task (to introduce the limits of the plank), ii) demonstration and evaluation of creative ideas (to introduce the hazard and observe safety voice/silence), and iii) wrap-up questionnaire (to ascertain safety concerns and study variables of interest). These stages are flanked by a participant welcome (to register participants and obtain informed consent) and debrief (to clarify the true study intent and ensure participants leave in the same psychological state they came in with). The study takes ±30-minutes per participant to complete.

To enable these stages, the protocol puts requirements on the general behaviour of research assistants, the laboratory setting and materials. The checklist provided at the end of this manual summarises these requirements. The pictures illustrate the environment and materials used to develop the paradigm.

General behaviour of research assistants

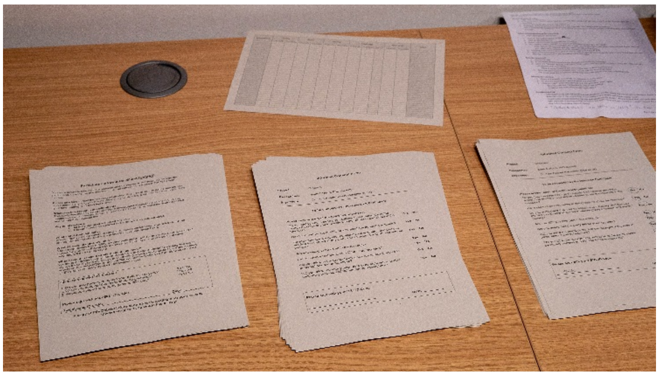
- During Welcome and Stages 1-3
    - Behaviours of the research assistant can have a strong influence on participant behaviours and can alter the outcomes of the study (e.g., positive, open, warm behaviours may encourage speaking up, and vice versa). Unless these behaviours form part of the study manipulations (e.g., leaders that display openness or closedness to participants' views), it is thus important to consider the following:
        - Be *neutral* and *factual* towards the participant (think 'be professional'), neither too friendly nor too distant.
    - Prompts can be made when participants do not follow through on instructions (e.g., making an idea).
    - If experimental manipulations put requirements on behaviour, act in line with conditions throughout the welcome and stages 1-3.
- During the Debrief
    - The debrief has two purposes: clarify the true study intention and make the participant feel well. Hence, during the debrief: always be friendly, warm, open to questions, etc.



Example of a reception area (entrance).

Laboratory environment

- A reception area for welcoming participants and paying out rewards
- Observation room with video recording facilities, or see-through mirror
- Quite experiment room(s)
- (optional) Provision of participant recruitment



Study materials

Participant reward

Materials

- 3 chairs
- A plank of wood (plywood, L: 120cm, W: 20cm, H:1.8cm)
- Blocks of wood (plywood, L: 3cm, W: 20cm, H:1.8cm)
- Table for the participant to sit at
- Questionnaire delivery method (i.e., iPad/PC, pen-and-paper).
- Duct tape
- Pens

Preparation

Reception Area

1) Provide seating for participants arriving early
2) Ensure that the observation equipment cannot be seen from the reception area

Observation room

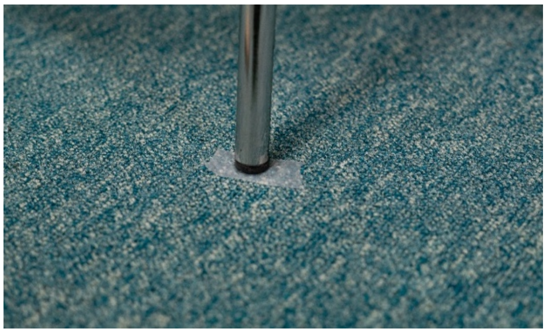
3) Start the video recording of the Experiment room(s) used.
4) Ensure that the following are placed at a logical location, so you are ready to pick these up when you need them:
   i) A few working pens.
   ii) Participant sign-up sheet (if laboratory provides advanced recruitment)
   iii) Study information (BRIEF) / Informed consent sheet
   iv) Study information (DEBRIEF)
   v) The Creativity-Feasibility form
   vi) Payment confirmation sheet (if laboratory requires receipts)
   vii) Money (i.e., participant reward)

Experiment room

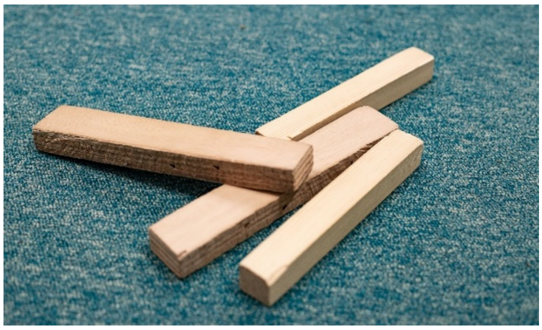
5) Set up:
   i) **Chairs:** Duct tape/mark two chairs to the ground at a distance on which the plank can be placed stably as a 'footbridge' between them.
   ii) **Plank of woods:** place the plank casually standing upright against one of the chairs.
   iii) **Blocks of wood:** place these casually at the base of the plank.
   iv) **Table & chair:** Place a single chair at the table for the participant to sit at while writing (no duct taping). The participant needs to face the plank and chairs.
   v) **iPad:** ensure a link has been added to favourites, so it can be easily re-started for the next participant.

Reception area (seating)

Observation Room equipment



Location of chairs discretely marked to the ground



Blocks of wood

Checklist

Requirements		**OK?**
Laboratory environment	A reception area for welcoming participants and paying out rewards	
	Observation room with video recording facilities, or see-through mirror	
	Quite experiment room(s)	
	Participant recruitment	
Key decisions	Manipulations: Which experimental manipulations are implemented and require tailoring of the protocol?	
	General behaviour of research assistants: how should research assistants behave in light of the experimental manipulations? (default = ‘professional’)	
Logistics	Ethical approval obtained	
	Laboratory space and time booked	
	Participant recruitment arranged	
	Research assistants allocated to timeslots	
Materials	A plank of wood (plywood, L: 120cm, W: 20cm, H:1.8cm)	
	Blocks of wood (plywood, L: 3cm, W: 20cm, H:1.8cm)	
	3 chairs	
	Table for the participant to sit at	
	Questionnaire delivery method (i.e., electronic, pen-and-paper).	
	Tape	
	Pens	
	Tracking sheet (made in Excel to track: participant IDs, timeslots, safety voice behaviours, condition allocation, notes)	
	Participant reward (e.g., money)	
Printing jobs (can be provided electronically)	Informed consent / brief form (printed in sufficient numbers)	
	Debrief form (printed in sufficient numbers)	
	Feasibility-creativity form for observing voice (printed in sufficient numbers)	
	Manual	
	Payment confirmation sheet (may come as laboratory facility)	
	Participant sign-up sheet (may come as laboratory facility)	

Protocol Steps

Stage 1: Participant Welcome

Dependent on the laboratory facilities, participants may enter the laboratory in varying ways. However, a reception area with seating is recommended.

1) Participant arrives in reception area
2) When the participant time slot is about to start:
   (1) Check whether the participant has arrived.
   (late participants should be denied participation, unless this can be accommodated)
   (2) Request participant ID code (if the laboratory utilises this) and note down attendance on the participant sheet.
3) Register the participant and acquire informed consent
   (1) Provide the 'Study Information (BRIEF) / Informed consent' sheet (see appendix D).
   (2) Say:

"Here is some high-level information about what to expect in the study. Please read this carefully and answer the questions on the back of the sheet".

---

   (3) Ensure the participant has agreed to all informed consent questions. (Participants are not allowed to participate without full agreement!).

Stage 2: 'Creativity' Task

1) Take the participant through to the experiment room.
2) Say:

"Please follow me."

---

3) Direct the participant to sit down.
4) Say:

"I am the research assistant for this study about creativity. The study involves 3 stages. First, you conduct a creativity task to develop creative ideas, then the feasibility and creativity of these ideas will be tested. Finally, there is a closing questionnaire."

---

5) Provide the Creativity Task and step back.
   (1) This can be either paper-based, or via an online questionnaire tool (e.g., Qualtrics, SurveyMonkey, etc.)
   (2) To dodge questions, say:

"The instructions make everything clear, please follow the instructions as provided."

---

6) Walk out of the experiment room for 5 minutes while the participant completes the task.
   (1) Keep track of the time
      (a) For iPad-based creativity tasks: these can be built to automatically transition (include a message to collect the research assistant) but make sure to keep track of participants as they may not collect the researcher.
      (b) For paper-and-pen-based creativity tasks: keep track of time yourself.

Stage 3: Demonstration and evaluation of Creative ideas

1) Take the 'Feasibility-Creativity form' with you.
2) Walk back into the Experiment room.
3) Explain the demonstration and evaluation of the creative ideas:
   (1) Say:

"Okay. The next stage involves testing these ideas for two things: feasibility, either a yes or no, and creativity on a scale of 1-5 with 5 being high. However, your ideas will be tested by the next participant, and the ideas of the previous participant are tested."

---

4) Go through the list of creative ideas from the 'previous participant'. For each idea:
   (1) Build: Let the participant build the idea.
   (2) Engage: Engage with the idea briefly (as if quickly evaluating what is built).
   (3) Feasibility: Ask whether the participant considers the idea feasible (request a: yes/no).
   (4) Creativity: Ask whether the participant considers the idea creative (request a: 1-5).
5) Upon encountering the '*footbridge'* idea.
   (1) Say:

"Hmm. That is actually pretty obvious, but I have not seen it before: could you please build a footbridge?"

---

   (2) Follow the same order: build, engage, feasibility, creativity.
      (a) If required, prompt participants to build a footbridge using the two chairs (unless they voice).
   (3) Before walking towards the plank, say:

"I will now to test whether this is a footbridge."

---

6) The research assistant walks up to the plank across the two chairs and walks over it.
7) Observe voice vs silence:
   (1) Note down on the form whether the participant speaks up about the plank being unsafe to walk on (before you have walked on it). For example, speaking up can sound like: "The plank stated a maximum weight", "That should be fine for a child (not you)"
   (2) Voice should only be coded if it occurs between two time-points:
      (a) After: the participant is asked to build the footbridge
      (b) Before: the RA is finished walking the plank and has stepped onto the ground.

Stage 4: Wrap-up Questionnaire

1) Finish demonstrating the creative ideas.
2) Say:

"For the final stage of the study, please complete this questionnaire."

---

3) Provide the wrap-up questionnaire
   (1) If paper-and-pen-base: provide questionnaire and pen.
   (2) If electronic: provide iPad/PC.
   Note: the electronic questionnaire can be built to automatically advanced from the Creativity Task during Stage 3, but ensure this takes sufficient time for the RA to not see the questionnaire prematurely.
4) Leave the room while the participant completes the questionnaire.

Stage 5: Debrief

Note: The intention of the debrief is to provide participants with a sense that they contributed to research, and explain to them that it was not about creativity, but speaking up about safety. This tends to work best when research assistants treat the debrief as an informal, open, and friendly conversation.

1) Take with you the Information sheet (DEBRIEF) and catch the participant before they come back to the waiting room and take them back to the experiment room.
2) Say:

"Thank you. This is the end of the study. However, I'd like to just take 2 minutes to debrief you on the study. I have to admit something: you may have guessed this study is actually not about creativity, but about how people speak-up and the factors that influence this."

---

3) Provide Participant Debrief Information Sheet. Say:

"Please take your time to read this and ask me any question you would have."

---

4) Discuss questions with the participants (or refer to the lead researcher).
5) Ensure the participant completes the debrief questions.
   (1) If a participant considers the research unethical: talk this through with them (they often misread the questions) and inform the lead researcher.
6) Take participants back to the reception area. Say:

Okay. Unless you have any remaining questions, let's go back and get you the reward for your time. Please follow me.

---

7) Pay the participant their reward.

Appendix D: Walking the Plank Materials

Participant Information Sheet (Brief)

You are being invited to take part in a research study. Before deciding to participate it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information about the study and don't feel rushed.

What is this research about?
You will engage in a task about creativity and associated factors.

Who is doing this research?
This study is led by [name], [role] at [department], [institution].

Why have you asked me to participate?
You have been recruited via [recruitment method].

What will participation involve?
- Participation involves a number of tasks around creativity
- Some participant may receive a closing questionnaire or interview
- A video recording may be made

How long will participation take?
Up to 30 minutes.

What about confidentiality?
Your data will be saved anonymously. Video recordings will be analysed by the experimenter and not be shared beyond the research team (unless you give explicit and voluntary approval after the study).

If you are willing to participate,
then please sign the Consent Form on the reverse of this page.
You can keep a copy of this Information Sheet for your records
(please enquire)

Informed Consent Form

Project: Study on Creativity

Researcher: [names]
Supervisor: [names]

To be completed by the Research Participant

Please answer each of the following questions:

Do you feel you have been given sufficient information about the research to enable you to decide whether or not to participate in the research?	**Yes**	**No**
Have you had an opportunity to ask questions about the research?	**Yes**	**No**
Do you understand that your participation is voluntary, and that you are free to withdraw at any time, without giving a reason, and without penalty?	**Yes**	**No**
Are you willing to take part in the research?	**Yes**	**No**
Are you aware that the study will be video recorded?	**Yes**	**No**
Will you allow the research team to use anonymized quotes in presentations and publications?	**Yes**	**No**
Will you allow the anonymized data to be archived, to enable secondary analysis and training future researchers?	**Yes**	**No**

The lab employs a no-name policy.
Please sign with your *ID Code*:

________________________________ **Date:** __________

Debrief form

You were invited to take part in a research study. However, to not invalidate findings (by participants knowing the true intention), the true nature of this research had to be kept hidden. Please take time to read the following information. Feel free to discuss issues with anyone, and if there is anything which is not clear or if you have any questions, feel free to ask. Take your time reading, and don't feel rushed.

What was this research *actually* about?

This study was not only about creativity. It investigated whether people speak-up about unsafe events. To promote safety, it investigated 'speaking up' about the 'footbridge' idea and factors that influence this.

Key to understand:

1. All materials used are tested and safe.
2. Research assistants played a role following an outlined script.

What about confidentiality?

This remains unchanged. We will treat your data video recording in strictest confidentiality. Your video will not be shared beyond the research team unless you give permission below.

What if I have changed my mind?

You are completely free to withdraw your participation without any consequence, or loss of reward for your time. Please indicate this to us if this case. Your answers and video recording will then be destroyed.

What if I want to raise a concern about this study?

We are open to listen to any concerns you have and are committed to working with you to resolve these (after all, this is what we study). If you want to raise any remaining issues, please contact the experimenter's supervisor ([email]), the [name laboratory] ([email]) or the [name institution] ethics committee ([email]).

Given this debrief information:

1. Do you give permission to use your data for analyses?	Yes / No
2. Would you consider this research to be *un*ethical?	Yes / No
3. Would you allow future participants to take part?	Yes / No

The lab employs a no-name policy.

Please sign with your ID code:

Participants ID code: ______________________ **Date:** __________

A copy of this Debrief Information Sheet is available for your records.
(please enquire if you would like a copy)

The creativity-feasibility form

Participant ID	Shelving		Mirror		Juggling		Footbridge			Piece of art		Notes
	Creativity	Feasibility	Creativity	Feasibility	Creativity	Feasibility	Creativity	Feasibility	Voice?	Creativity	Feasibility	

Appendix E: Jupyter notebook for Article 3

Table of Contents

Introduction to notebook

This notebook presents the full statistical analyses as described in the accompanying article. You are recommended to read the article alongside this file to contextualise the analyses and results.

The relationship between the article and this notebook is as such that analyses in the section 'Results' are mapped onto the results within the article. The analyses follow the order they are presented within the article, unless specified otherwise. The notebook further describes which variables are used and selected for the analyses, but does not provide an in-depth interpretation of the findings. This is provided within the accompanying article.

Please note: the required datafile is provided alongside the notebook in the supplementary materials of the article. The reliability analysis for coding videos and the word vector analysis are not contained in this file.

Load

Modules

In [ ]:

```
# LOADING MODULES
import pandas as pd
import numpy as np
import pathlib
import seaborn as sns
import re
import os
import json
import scipy
from pathlib import Path
import scipy.stats as stats
from scipy.stats import pointbiserialr
import pandas_profiling
from statsmodels.stats.multicomp import pairwise_tukeyhsd
import statsmodels.api as sm
```

```
import statsmodels.api as sm
from statsmodels.formula.api import ols

# SPACY
import spacy
nlp = spacy.load('en_core_web_sm')
from spacy.matcher import Matcher
from spacy import displacy
import scattertext as st

# SETTING UP THE NOTEBOOK
from IPython.display import IFrame
from IPython.core.display import display, HTML
display(HTML("<style>.container { width:75% !important; }</style>"))
import matplotlib.pyplot as plt
%matplotlib inline
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"
```

Custom functions

In [ ]:

```
def correlationTable(df, method='spearman', decimals=3, missing_values=False):
    '''
    inputs: df
            method='pearson' or 'spearman'
            decimals = for tables
            missing_values= print table of missing values
    output: table of correlations
    '''

    from scipy.stats import pearsonr, spearmanr

    # ONLY USE NUMERIC VARIABLES
    df_not_numbers = df.select_dtypes(exclude=[np.number])
    if len(df_not_numbers)>0:
        print('Non numeric columns: ', [x for x in df_not_numbers.columns])
        print('If these strings are binary variables, spearman method is used')
    df_test = df.copy()

    # CHECK MISSING VALUES
    for var in df_test.columns:
        pct_missing = int((sum(df_test[var].isna())/len(df_test.index))*100)
        if pct_missing > 30:
            print(f'{pct_missing} missing data for {var}')

    ## CALCULATING CORR, P-VALUE, & N
    df_testcols = pd.DataFrame(columns=df_test.columns)
    corr_table = df_testcols.transpose().join(df_testcols, how='outer')
    for r in df_test.columns:
        for c in df_test.columns:
            x = df_test[r]
            y = df_test[c]
            df_nona = pd.DataFrame({'x': x, 'y': y})
            df_nona = df_nona.dropna().copy()
            # get correlations
            if method=='spearman':
                corr_table[r][c] = spearmanr(df_nona['x'], df_nona['y'])
            if method=='pearson':
                corr_table[r][c] = pearsonr(df_nona['x'], df_nona['y'])
            # add number of data points for each correlation
            corr_table[r][c] = corr_table[r][c]+ (df_nona.shape[0],)
    # table of correlations
    c_table =  corr_table.apply(lambda x: [y[0] for y in x]).round(decimals)
    # table of p values
    p_table = corr_table.apply(lambda x: [y[1] for y in x])
    # table of n
    n_table = corr_table.apply(lambda x: [y[2] for y in x])
    ## P-VALUE STARS
    # create three masks
    r1 = c_table.applymap(lambda x: '{}*'.format(x))
    r2 = c_table.applymap(lambda x: '{}**'.format(x))
    r3 = c_table.applymap(lambda x: '{}***'.format(x))
    # apply them where appropriate
```

```
    p_values = [0.05, 0.01, 0.001]
    c_table = c_table.mask(p_table<=p_values[0],r1)
    c_table = c_table.mask(p_table<=p_values[1],r2)
    c_table = c_table.mask(p_table<=p_values[2],r3)
    ## STYLING
    def color_sigs(val):
        if '*' in str(val):
            if '-' in str(val):
                color = 'blue'
            else:
                color = 'red'
        else:
            color = 'black'
        return 'color: %s' % color
    c_table = c_table.style.set_properties(**{'text-align': 'center'}).applymap(color_sigs)
    if missing_values==True:
        # '_r' reverses color map
        n_table = n_table.style.background_gradient(cmap='Reds_r', axis=None)
        display(n_table)
    display(c_table)
    print(f'* p<{p_values[0]};  ** p<{p_values[1]};  *** p<{p_values[2]}')
    return
```

In [ ]:

```
def loadPatterns(path_to_patterns_folder, file_search_string, label_search_string=''):
    """
    Loads pattern files from folder that contain search string.
    input:  path_to_patterns_folder = string (e.g., '.\patterns')
            file_search_string = (accepts '|'/or), to find file(s) in folder
            label_search_string =(accepts '|'), to find label(s) in file
    output: matcher has patterns loaded
    """

    # reset matcher and make global
    global matcher
    matcher = Matcher(nlp.vocab)

    # LOAD PATTERNS
    files_in_folder = os.listdir(pathlib.Path(path_to_patterns_folder))
    files_in_folder = [f for f in files_in_folder if '.jsonl' in f]
    search_list = file_search_string.split('|')
    pattern_files = []
    for search_pattern in search_list:
        for file in files_in_folder:
            if search_pattern in file:
                pattern_files.append(file)
    patterns = []
    for file in pattern_files: # load multiple files
        pattern_file = pathlib.Path(path_to_patterns_folder+'//'+file)
        with pattern_file.open() as f:
            line_no = 0
            for line in f:
                line_no = line_no + 1
                try:
                    if line[0] == '{':
                        data = json.loads(line)
                        patterns.append(data)
                    else:
                        pass
                except:
                    print(f'PATTERN ERROR\nFILE: {pattern_file}\nLINE NUMBER: {line_no}\n{line}')
                    return

    # FILTER PATTERNS BY LABEL_SEARCH_STRING
    label_search_strings = label_search_string.split('|')
    patterns_filtered = []
    for s in label_search_strings:
        for p in patterns:
            if s in p['LABEL']:
                patterns_filtered.append(p)

    # ADD PATTERNS TO MATCHER
    labels = list(set([x['LABEL'] for x in patterns_filtered]))
    for label in labels:
        # check label is in the vocab
```

```
        if label not in nlp.vocab: # if the label is not in the vocab
            lex = nlp.vocab[label]# then add the label to the vocab
            assert label in nlp.vocab, f'tried to add {label} to nlp.vocab, but failed!'

        # add list of patterns to the matcher
        label_patterns = [x['PATTERN'] for x in patterns if x['LABEL'] == label]
        matcher.add(label, None, *label_patterns)# the '*' is needed for list of patterns
```

In [ ]:

```
def viewMatchesDocs(df, doc_column, id_column, path_to_patterns_folder, file_search_string,
label_search_string = '', max_matches=40):
    """
    Loads spacy pattern file and displays matches in series of docs
    input:  df = pandas dataframe
            doc_column = name of column of spacy docs in dataframe
            id_column = id to print with matches,
            path_to_patterns_folder = string (e.g., '.\patterns')
            file_search_string = (accepts '|'/or), to find file(s) in folder
            label_search_string =(accepts '|'), to find label(s) in file
    output: display matches
    """

    # Load patterns
    loadPatterns(path_to_patterns_folder, file_search_string, label_search_string)

    # Create df
    df_search = df[[id_column]+[doc_column]]
    print('\nNumber of docs: ', len(df))

    # ITERATE OVER DOCS
    start_old = 0
    match_counter = 0
    doc_counter = -1
    for index, row in df_search.iterrows():
        doc_id = str(row[id_column])
        doc=row[doc_column]
        matches = matcher(doc)
        doc_counter = doc_counter + 1
        if len(matches)>0:
            print(f'\nDoc number: {doc_counter}\n{repr(id_column)}: {doc_id}\nWordcount: {len(doc)}
')
            match_counter = match_counter+1
            if match_counter > max_matches:
                print(f'\n\nmax_matches ({max_matches}) reached')
                return

            # CREATE DISPLACY MATCHES (use character position, not word position)
            displacy_matches = []
            match_ents = []
            match_labels = []
            for match in matches:
                match_id, start, end = match
                match_label = nlp.vocab.strings[match_id]
                match_labels.append(match_label)
                span = doc[start:end]
                match_ent = {'start': span.start_char,
                             'end': span.end_char,
                             'label': nlp.vocab.strings[match_id]}
                match_ents.append(match_ent)

            # PRINT LABELS
            labels = list(set(match_labels))
            labels.sort()

            displacy_matches.append({'text': doc.text, 'ents': match_ents, 'title': None,
'settings':{}})
            displacy.render(displacy_matches, style='ent', jupyter=True, manual=True)
        else:
            pass
```

In [ ]:

```
def logisticR(df, predictors, outcome, odds_ratio=False):
```

```
    df_test = df[[outcome]+predictors]
    display(df_test.head(2))
    X = df_test.copy()
    y = df_test[outcome].copy()
    X.drop([outcome], axis=1, inplace=True)
    X = sm.add_constant(X)
    model = sm.Logit(y, X.astype(float))
    result = model.fit()
    display(result.summary())
    if odds_ratio==True:
        # ODDS RATIOS
        print('ODDS RATIOS')
        conf = result.conf_int() # confidence intervals
        conf['OR'] = result.params
        conf.columns = ['2.5%', '97.5%', 'Odds Ratio']
        display(np.exp(conf))
```

In [ ]:

```
def multLinR(df, predictors, outcome):
    df_test = df[[outcome]+predictors]
    display(df_test.head(2))
    X = df_test.copy()
    y = df_test[outcome].copy()
    X.drop([outcome], axis=1, inplace=True)
    X = sm.add_constant(X)
    model = sm.regression.linear_model.OLS(y, X.astype(float))
    result = model.fit()
    display(result.summary())
```

In [ ]:

```
def moderationAnalysis(df, x_interven, x_moder, y_cont):
    '''
    input:  df
            x_interven = intervention variable
            x_moder = moderator variable (binary)
            y_cont = continuous variable
    '''
    from io import StringIO

    print('DESCRIPTIVES')
    display(df[[x_interven, x_moder, y_cont]].groupby([x_interven, x_moder]).agg([len, np.mean, np.
std]).round(3))

    print('MODEL SUMMARY')
    formula_string = y_cont+' ~ '+x_interven+'*'+x_moder
    model = ols(formula_string, data=df).fit()
    display(model.summary())

    print(f'CONDITIONAL EFFECT OF {x_interven} ON {y_cont} AT VALUES OF {x_moder}')
    def extractSummaryRow(summary, variable):
        csv = summary.tables[1].as_csv()
        df_new = pd.read_csv(StringIO(csv), index_col=0)
        df_new.index = df_new.index.str.strip()
        return df_new.loc[[variable]]
    row0 = extractSummaryRow(model.summary(), x_interven)
    row0.rename(index={x_interven: '0.0'}, inplace=True)
    df_adjusted = df.copy()
    df_adjusted[x_moder] = 1 - df_adjusted[x_moder]
    model_adjusted = ols(formula_string, data=df_adjusted).fit()
    row1 = extractSummaryRow(model_adjusted.summary(), x_interven)
    row1.rename(index={x_interven: '1.0'}, inplace=True)
    table = pd.concat([row0, row1])
    table.index.name = x_moder
    table.rename(columns=lambda x: x.strip(), inplace=True)
    table.rename(columns={'coef': 'effect'}, inplace=True)
    display(table)
```

In [ ]:

```
def anova(df, x_cat, y_cont, tukey=False, power_analysis=True, plot=False):
    '''
    input:  df
```

```
            x_cat = categorical variable
            y_cont = continuous variable
    output: anova, plot, power
    '''

    print('DESCRIPTIVES')
    display(df[[x_cat, y_cont]].groupby(x_cat).agg([len, np.mean, np.std]).round(3))
    print('ANOVA')
    formula_string = y_cont+' ~ C('+x_cat+')'
    model = ols(formula_string, data=df).fit()
    # ANOVA TABLE
    anovaTable = sm.stats.anova_lm(model, type=2)
    anovaTable['mean_sq'] = anovaTable[:]['sum_sq']/anovaTable[:]['df']
    anovaTable['eta_sq'] = anovaTable[:-1]['sum_sq']/sum(anovaTable['sum_sq'])
    anovaTable['omega_sq'] = (anovaTable[:-1]['sum_sq']-(anovaTable[:-1]['df']*anovaTable['mean_sq'
][-1]))/(sum(anovaTable['sum_sq'])+anovaTable['mean_sq'][-1])
    cols = ['sum_sq', 'df', 'mean_sq', 'F', 'PR(>F)', 'eta_sq', 'omega_sq']
    anovaTable = anovaTable[cols]
    display(anovaTable.round(3))
    if tukey==True:
        print('TUKEY POST-HOC')
        print(pairwise_tukeyhsd(endog=df[y_cont], groups=df[x_cat], alpha=0.05))

    # PLOT
    if plot==True:
        plot_width = len(set(df[x_cat]))*2
        sns.set(rc={'figure.figsize':(plot_width,4)})
        sns.violinplot(x=x_cat, y=y_cont, data=df, scale='count')
```

In [ ]:

```
from numpy.linalg import eigvals, inv, solve


def s_values(mdl, tol=1e-8):
    '''
    input:  mdl = MANOVA model
            tol = eigenvalue tolerance (smaller than this is considered 0)
    output: ranks of sums of squares and products matrices for each component
            of MANOVA model
    '''
    # Set up contrast matrices for each component hypothesis
    terms = mdl.data.design_info.term_name_slices
    hypotheses = []
    for key in terms:
        L_contrast = np.eye(mdl.exog.shape[1])[terms[key], :]
        hypotheses.append([key, L_contrast])

    # Function for producing sums of squares and products (SSP) matrices
    def fn(L, M):
        params, df_resid, inv_cov, sscpr = mdl._fittedmod
        t1 = L.dot(params).dot(M)
        t2 = L.dot(inv_cov).dot(L.T)
        H = t1.T.dot(inv(t2)).dot(t1)
        E = M.T.dot(sscpr).dot(M)
        return E, H

    # Calculate SSP matrices for each hypothesis and determine rank
    # from eigenvalues
    k_yvar = len(mdl.endog_names)
    results = {}
    for hypo in hypotheses:
        name, L = hypo
        M = np.eye(k_yvar)
        E, H = fn(L, M)
        EH = np.add(E, H)
        eigv2 = np.sort(eigvals(solve(EH, H)))
        s = (eigv2 > tol).sum()
        results[name] = s
    return results
```

In [ ]:

```
def manova(df, x_cat, y_cont):
    '''
```

```
    input:  df
            x_cat = list of categorical variables
            y_cont = list of continuous variables
    output: descriptive summary, manova, anovas
    '''

    print('DESCRIPTIVES')
    def count(x):
        return (~np.isnan(x)).sum()
    display(df[x_cat+y_cont].groupby(x_cat).agg([count, np.mean, np.std]).round(3))

    print('MANOVA (MULTIVARIATE TESTS)')
    formula_lhs = '+'.join(y_cont)
    formula_rhs = '*'.join(['C(' + var + ')' for var in x_cat])
    formula = formula_lhs + ' ~ ' + formula_rhs
    mdl = sm.MANOVA.from_formula(formula, data=df)
    res = mdl.mv_test()
    s_dict = s_values(mdl)
    smry = res.summary_frame.copy()
    # Partial eta squared functions
    pes_fns = {"Wilks' lambda": lambda x, s: 1 - x**(1/s),
               "Pillai's trace": lambda x, s: x/s,
               "Hotelling-Lawley trace": lambda x, s: x/(x+s),
               "Roy's greatest root": lambda x, s: x/(x+1)}
    pes_values = pd.Series(index = smry.index.copy())
    for row in smry.itertuples():
        ix = row.Index
        name, stat = ix
        f = pes_fns[stat]
        pes_values.loc[ix] = f(row.Value, s_dict[name])
    smry['Partial Eta Sq'] = pes_values
    smry = smry.astype(np.float64).round(3)
    display(smry)

    print('ANOVAS (TESTS OF BETWEEN-SUBJECT EFFECTS)')
    def oneAnovaTable(df_local, x_list, y):
        # Fit ANOVA model
        rhs = '*'.join(['C(' + var + ')' for var in x_list])
        model = ols(y + ' ~ ' + rhs, data=df_local).fit()
        anovaTable = sm.stats.anova_lm(model, type=2)
        # Add columns to ANOVA table
        anovaTable['mean_sq'] = anovaTable[:]['sum_sq']/anovaTable[:]['df']
        anovaTable['eta_sq'] = anovaTable[:-1]['sum_sq']/sum(anovaTable['sum_sq'])
        pes = pd.Series(index = anovaTable.index.copy())
        ss_resid = anovaTable['sum_sq'].loc['Residual']
        for row in anovaTable.itertuples():
            if row.Index != 'Residual':
                pes.loc[row.Index] = row.sum_sq / (row.sum_sq + ss_resid)
        anovaTable['partial_eta_sq'] = pes
        # Adjust indexing for later output appearance including sorting
        anovaTable.index.name = 'source'
        anovaTable['dep_var'] = y
        anovaTable.set_index('dep_var', append=True, inplace=True)
        anovaTable['temp'] = range(anovaTable.shape[0])
        anovaTable.set_index('temp', append=True, inplace=True)
        cols = ['sum_sq', 'df', 'mean_sq', 'F', 'PR(>F)', 'eta_sq', 'partial_eta_sq']
        return anovaTable[cols]
    # Combine tables for each dependent variable
    tables = [oneAnovaTable(df, x_cat, y) for y in y_cont]
    combined = pd.concat(tables)
    combined.sort_index(level=2, inplace=True)
    combined.index = combined.index.droplevel(level=2)
    display(combined.round(3))
```

Data

In [ ]:

```
# Please specify the filename and location for your local device.
file = "Supplementary file - data.csv"
df_raw = pd.read_csv(file, na_values=[-999, '#NULL!'])
df_raw.shape
# check dataframe
with pd.option_context('display.max_rows', 4, 'display.max_columns', None):
```

```
    display(df_raw)
```

Data cleaning and variable selection

- Safety voice: safetyvoice_binary (1= voice)
- Concerns: concerned_binary (1 = concerned)
- Concern strength: concerned (1 = strongly disagree)
- Hazard salience: condition_awareness (1 = salient, 0 = control)
- Encouragements: condition_discourage (1 = discouraged, 0 = encouraged)
- Responsibility: condition_responsibility (1 = clear, 0 = diffuse)
- Hazard stages: voice_stage_v3 (0:'silence', 1:'concept', 2:'encounter', 3:'hazard', 4:'posthoc')

In [ ]:

```
#make variable names lower case
df_raw.columns = [col.lower() for col in df_raw.columns]
# make variable text values lower case
df_raw['safetyvoice_binary'] = [str(val).lower() for val in df_raw['safetyvoice_binary']]
```

There are 10 participants without recorded videos and 2 that did not answer whether they were concerned (1 overlaps). -> Drop n = 12.

In [ ]:

```
df_raw.shape #Raw
df_raw = df_raw[df_raw.data_for_analysis != 0].copy() #removes missing observations for video and
concerned.
df_raw.shape
```

Map values for the categorical variables.

In [ ]:

```
# CREATE STRING VERSIONS OF BINARY VARIABLES
df_raw['condition_awareness_string'] = df_raw['condition_awareness'].map({0:'no', 1:'yes', }).astyp
e(str)
df_raw['condition_discourage_string'] = df_raw['condition_discourage'].map({0:'no', 1:'yes', }).ast
ype(str)
df_raw['condition_responsibility_string'] = df_raw['condition_responsibility'].map({0:'no', 1:'yes'
}).astype(str)
df_raw['concerned_binary_string'] = df_raw['concerned_binary'].map({0:'no', 1:'yes', }).astype(str)
df_raw['safetyvoice_binary'] = df_raw['safetyvoice_binary'].map({'0.0':0.0, '1.0':1.0, })
df_raw['safetyvoice_binary_string'] = df_raw['safetyvoice_binary'].map({0.0:'no', 1.0:'yes', })
df_raw['safetyvoice_binary_string'].value_counts()
```

In [ ]:

```
df_raw['voice_stage'] = df_raw['voice_stage'].map({0:'silence', 1:'concept', 2:'encounter', 3:'haza
rd', 4:'posthoc'}).astype(str)
df_raw['voice_stage'].value_counts()
```

In [ ]:

```
df_raw['typology_group'] = df_raw['typology_group'].map({1:'voice & concern', 2:'voice & no concern
', 3:'silence & concern', 4:'silence & no concern'}).astype(str)
df_raw['typology_group'].value_counts()
```

Calculate text variables

In [ ]:

```
# Create temporary back-up.
df0 = df_raw.copy()
```

Extract participant text

In [ ]:

```
# creates separate variables for researcher and participant text

def getResearcherText(sentences_all):
    sents = [sent for sent in sentences_all if sent.strip().startswith('R:')]
    sents = [sent.replace('R:', '').strip() for sent in sents]
    joined = ' '.join(sents)
    return joined

def getParticipantText(sentences_all):
    sents = [sent for sent in sentences_all if sent.strip().startswith('P:')]
    sents = [sent.replace('P:', '').strip() for sent in sents]
    joined = ' '.join(sents)
    return joined

df0['text_researcher_cleaned'] = df0['transcribed_text'].astype(str).str.split('\n').apply(getResea
rcherText)
df0['text_participant_cleaned'] = df0['transcribed_text'].astype(str).str.split('\n').apply(getPart
icipantText)
```

In [ ]:

```
# removes all text in square brackets

import re
regex_search = re.compile(r"\[[a-z :]*\]")#searches for 'space[any a-z or space characters]'

def removeBracketedText(original_text):
    cleaned_text = re.sub(regex_search, '', original_text)
    return cleaned_text

df0['text_researcher_cleaned'] = df0['text_researcher_cleaned'].apply(removeBracketedText)
df0['text_participant_cleaned'] = df0['text_participant_cleaned'].apply(removeBracketedText)
```

In [ ]:

```
# convert text into spacy doc for nlp
df0['text_researcher'] = df0['text_researcher_cleaned'].apply(nlp)
df0['text_participant'] = df0['text_participant_cleaned'].apply(nlp)
```

Add wordcounts

In [ ]:

```
df0['wordcount_researcher'] = df0['text_researcher'].apply(len)
df0['wordcount_participant'] = df0['text_participant'].apply(len)
```

In [ ]:

```
df0.hist(['wordcount_researcher', 'wordcount_participant'], bins=40, figsize=(12, 3));
```

The Safety voice Typology: Voice/Silence vs Concern/No Concern

This plot uses scaled f scores instead of raw scores, so that the voice/silence and concern/no concern axes are comparable. It displays the 2 by 2 typology for (un)concerned voice and silence (Noort et al., 2019).

In [ ]:

```
# Please note: This graph is created, saved externally and then loaded as a seperate file into the
notebook. Accordingly, it is not shown in .html versions of the notebook.

corpus_voice = st.CorpusFromParsedDocuments(df0, category_col='safetyvoice_binary_string',
parsed_col='text_participant').build()
corpus_concern = st.CorpusFromParsedDocuments(df0, category_col='concerned_binary_string',
```

```
corpus_concern = st.CorpusFromParsedDocuments(df0, category_col='concerned_binary_string',
parsed_col='text_participant').build()

concern_f_scores = corpus_concern.get_scaled_f_scores('yes')
voice_f_scores = corpus_voice.get_scaled_f_scores('yes')

html = st.produce_scattertext_explorer(corpus_voice,
                                       category='yes',
                                       category_name='Voice',
                                       not_category_name = 'Not Voice',
                                       width_in_pixels=1000,
                                       scores=voice_f_scores,
                                        x_coords=concern_f_scores,
                                        x_label='More Concern',
                                        y_coords=voice_f_scores,
                                        y_label='More Voice',
                                       show_characteristic=False,
                                        show_top_terms=True,
                                       minimum_term_frequency=5,
                                       use_full_doc=True,
                                       p_value_colors=True,
                                       max_snippets=None
                                       )
file_name = 'scatter_concernVoice.html'
open(file_name, 'wb').write(html.encode('utf-8'))
IFrame(src=file_name, width=1300, height=900)
```

View dictionaries

In [ ]:

```
# Place a folder named 'dictionaries' in the same folder as your notebook.
path_to_patterns = r"./dictionaries"

for x in os.listdir(pathlib.Path(path_to_patterns)):
    print(x)
```

In [ ]:

```
## VIEW PATTERNS IN CONTEXT
viewMatchesDocs(df0, doc_column='text_participant', # text to search
                id_column='id_final',
                path_to_patterns_folder=path_to_patterns,
                file_search_string='custom',      # pattern file to load
                label_search_string='custom_dict_cautionary_wv', # patterns to show as example
                max_matches=5)
```

In [ ]:

```
viewMatchesDocs(df0, doc_column='text_participant', # text to search
                id_column='id_final',
                path_to_patterns_folder=path_to_patterns,
                file_search_string='with_vectors',      # pattern file to load
                label_search_string='oblique', # patterns to show
                max_matches=5)
```

Create variables from dictionaries

The dictionaries for safety concerns, safety voice and the LIWC dictionaries are loaded here. The LIWC dictionaries are propriety and are therefore not provided.

In [ ]:

```
# LOAD ALL DICTIONARIES TOGETHER
path_to_patterns = r"./dictionaries" # the folder to search for patterns
loadPatterns(path_to_patterns, 'liwc|custom|Vagueness')# a search term to find pattern files

# FIND MATCHES
df0['pattern_matches'] = df0['text_participant'].apply(matcher)
```

In [ ]:

```
# FUNCTION TO COUNT MATCHES
# takes match object; returns dictionary of counts of matches
# if the count is zero, then no dictionary entry is returned (thus 'fillna' when making df0)
def scoreMatches(matches, doc, print_scores=False):
    matches_score = {}
    from collections import Counter
    counts = Counter(element[0] for element in matches)
    #print(counts)
    for entry in counts:
        # error handling for unable to find entry for label
        try:
            value = counts[entry]
            label = nlp.vocab.strings[entry]
            if print_scores == True:
                print('{:<20}: {:<10}  {:}%'.format(label, value))
            matches_score[label]=value
        except:
            print('EROR: ', entry, '\n', doc, '\n')
    return matches_score
```

In [ ]:

```
# count matches
df0['pattern_freq'] = df0.apply(lambda x: scoreMatches(x['pattern_matches'], x['text_participant'])
, axis=1)
# extract counts
df0 = pd.concat([df0, df0['pattern_freq'].apply(pd.Series).fillna(0)], axis=1, join='outer', sort=F
alse)
# create total vagueness score
df0['cvs_total'] = df0.loc[:, [col for col in df0.columns if 'cvs_' in col]].sum(axis=1)
```

Check new variables

In [ ]:

```
# view all variables
with pd.option_context('display.max_rows', 1, 'display.max_columns', None):
    display(df0)
```

In [ ]:

```
df = df0.copy()
```

Results

The cells below contain the results of the analyses. The section numbering refers to section numbers within the article.

3.1. Manipulation and dictionary checks

For the hypotheses to be tested, four checks are important and tested below:

1. The **scenario** should elicit safety concerns in general, (concerned) participants should speak-up, and participants should speak.
2. The **hazard salience manipulation** should elicit more, and stronger, safety concerns.
3. The **responsibility manipulation** should elicit a stronger felt responsibility.
4. The **encouragement manipulation** should elicit perceived social risk and imcrease participant wordcount.

Table 4 provided descriptives and correlations:

1. Descriptives and correlations.

Scenario

In [ ]:

```
# 1. Scenario (concerns and voice/silence)
dfconcerned = df[df.concerned_binary.isin(["1"])]
dfconcerned['concerned_binary'].value_counts()

stats.ttest_1samp(df.concerned_binary, 0) #Are people concerned?
stats.ttest_1samp(df.safetyvoice_binary, 0) #Do people speak-up?
stats.ttest_1samp(dfconcerned.safetyvoice_binary, 0) #Do concerned people speak-up?

# For providing numbers to Table 4.
##Frequencies safety voice variables.
df['safetyvoice_binary'].value_counts()
df['concerned_binary'].value_counts()
#Frequencies typology group
df['typology_group'].value_counts()
```

In [ ]:

```
# 1. Scenario (word counts)
for corpus in df[['wordcount_researcher', 'wordcount_participant']]:
    print('{:<30} total words = {:>20,}'.format(corpus, int(df[corpus].sum())))

anova(df, 'safetyvoice_binary','wordcount_participant')
```

Manipulations

In [ ]:

```
# 2. Hazard salience manipulation
anova(df, 'condition_awareness', 'concerned')

outcome = 'concerned_binary'
predictors = ['condition_awareness']

logisticR(df, predictors, outcome, odds_ratio=True)
```

In [ ]:

```
# 3. Responsibility manipulation
manova(df, ['condition_responsibility'], ['q_2', 'liwc_we'])
```

In [ ]:

```
# 4. Encouragement manipulation: additional variables.
df['condition_discourage'].value_counts(dropna=False)
manova(df, ['condition_discourage'],
['social_risk','wordcount_participant','liwc_informal','liwc_we','liwc_i','liwc_negate']) #Do
participants perceive more social risk?
```

Table 4 (means and correlations)

In [ ]:

```
# 5. Data for Table 4 (conditions).
df['condition_awareness'].value_counts()
df['condition_discourage'].value_counts()
df['condition_responsibility'].value_counts()
```

In [ ]:

```
vars_to_describe = [
'condition_awareness',
'condition_responsibility',
'condition_discourage',
'safetyvoice_binary',
'voice_reported',
'custom_dict_voicecomposite_wv',
'concerned',
```

```
'concerned_binary',
'felt_responsibility',
'social_risk',
'custom_dict_concerned_wv',
'custom_dict_disfluency_wv',
'custom_dict_informative_wv',
'custom_dict_inquisitive_wv',
'custom_dict_prohibitive_wv',
'custom_dict_cautionary_wv',
'custom_dict_oblique_wv',
'wordcount_researcher',
'wordcount_participant',
]
```

In [ ]:

```
# 5. Data for Table 4 (basic descriptives).
df[vars_to_describe].describe().T
```

In [ ]:

```
# 5. Data for Table 4 (detailed descriptives).
pandas_profiling.ProfileReport(df[vars_to_describe])
```

In [ ]:

```
# 5. Data for Table 4 (correlations).
## Spearman rank correlations are used for Table 4 due to skewed variables.
correlationTable(df[vars_to_describe], method='spearman')
```

Dictionary validation

The following dictionaries were validated:

Concerned dictionaries

1. Concerned dictionary (should be higher for participants self-reporting to be concerned)
2. Disfluencies dictionary (should be higher for participants self-reporting to be concerned)

Safety voice dictionaries

1. Informative safety voice
2. Inquisitive safety voice
3. Prohibitive safety voice
4. Cautionary safety voice
5. Oblique safety voice
6. Safety voice dictionary (composite dictionary)

(note: each safety voice dictionary should be higher for participants observed as engaging in the behaviour).

Correlations

1. Safety voice dictionary <-> self-reported safety voice
2. Safety voice measures <-> concern measures

In [ ]:

```
# 1. Concerned dictionary
# 2. Disfluencies dictionaries
manova(df, ['concerned_binary'], ['custom_dict_concerned_wv', 'custom_dict_disfluency_wv'])
```

In [ ]:

```
# (4-7 are commented out for succinctness but can be performed through removing the '#').

# 3. Informative safety voice
anova(df, 'informative','custom_dict_informative_wv')

# 4. Inquisitive safety voice
```

```
# 4. Inquisitive safety voice
#anova(df, 'inquisitive','custom_dict_inquisitive_wv')

# 5. Prohibitive safety voice
#anova(df, 'prohibitive','custom_dict_prohibitive_wv')

# 6. Cautionary safety voice
#anova(df, 'cautionary','custom_dict_cautionary_wv')

# 7. Oblique safety voice
#anova(df, 'oblique','custom_dict_oblique_wv')
```

In [ ]:

```
# 8. Safety voice dictionary
anova(df, 'safetyvoice_binary','custom_dict_voicecomposite_wv')
```

In [ ]:

```
# 9. Safety voice dictionary <-> self-reported safety voice
# 10. Safety voice measures <-> concern measures

## For obtaining the specific p-values of voice measure correlations:
print('Concern Dictionary - Concern strength report')
stats.spearmanr(df.custom_dict_concerned_wv,df.concerned)

print('Voice Dictionary - Concern dictionary')
stats.spearmanr(df.custom_dict_voicecomposite_wv,df.custom_dict_concerned_wv)
print('Voice Dictionary - Concern strength report')
stats.spearmanr(df.custom_dict_voicecomposite_wv,df.concerned)
```

3.2. Scaling safety silence

To test the extent to which safety silence can be scaled upon safety voice speech, we tested the following:

1. Overall wordcount: Do participants in the safety silence group (concerned & silence) score zero on their wordcount?
2. Thematic differences on LIWC dicitonaries: risk, concerns, perceptions
3. Safety voice wordcounts: Do participants in the safety silence group (concerned & silence) score zero on the safety voice dictionaries?
4. Are differences in safety voice dictionary scores are matter of degree?

The results below indicate that participants that do not speak-up about safety are not silent and, in fact, uttter words reflective of raising safety concerns. This involved participants silent about safety uttering informative, inquisitive, prohibitive, cautionary and oblique words (and words on the composite measure). Instead, the distinction between raising and withholding safety concerns is relevant, but as a matter of degree. That is, participants raising safety concerns utter more safety voice speech than those witholding concerns.

In [ ]:

```
# Preparation:

# Defining key variables to describe for analysing hypothesis 1a-e.
vars_to_describe_2 = [
'wordcount_participant',
    'custom_dict_voicecomposite_wv',
    'custom_dict_informative_wv',
    'custom_dict_inquisitive_wv',
    'custom_dict_prohibitive_wv',
    'custom_dict_cautionary_wv',
    'custom_dict_oblique_wv']

# Calculating the necessary dataframes
dfsilent = df[df.safetyvoice_binary.isin(["0"])]
dfvoice = df[df.voice_stage.isin(["hazard", "concept", "encounter"])]
dfsafsilent = df[df.typology_group.isin(["silence & concern"])]

dfvoice.shape
dfsilent.shape
dfsafsilent.shape

# Descriptives table for dictionary scores.
```

```
print('Dictionary scores for participants engaging in Safety Silence')
dfsafsilent[vars_to_describe_2].describe().T
```

In [ ]:

```
# 1. Overall wordcount.
## This test (similar to below) compares the scores to the absence of speech, which has the test-v
alue '0'.
print('*Participants that do not speak-up about safety are not silent*')
stats.ttest_1samp(dfsafsilent.wordcount_participant, 0)
```

In [ ]:

```
# 3. Safety voice dictionary scores for safety silence
print('*They utter safety voice words (composite)*')
stats.ttest_1samp(dfsafsilent.custom_dict_voicecomposite_wv, 0)

print('*They utter Informative words*')
stats.ttest_1samp(dfsafsilent.custom_dict_informative_wv, 0)

print('*They utter Inquisitive words*')
stats.ttest_1samp(dfsafsilent.custom_dict_inquisitive_wv, 0)

print('*They utter Prohibitive words*')
stats.ttest_1samp(dfsafsilent.custom_dict_prohibitive_wv, 0)

print('*They utter Cautionary words*')
stats.ttest_1samp(dfsafsilent.custom_dict_cautionary_wv, 0)

print('*They utter Oblique words*')
stats.ttest_1samp(dfsafsilent.custom_dict_oblique_wv, 0)
```

In [ ]:

```
# 1. Overall wordcount (cont'd).
# 2. Thematic differences on LIWC dicitonaries.
# 4. A matter of degree?
manova(dfconcerned, ['safetyvoice_binary'], ['wordcount_participant','custom_dict_informative_wv',
'custom_dict_inquisitive_wv', 'custom_dict_prohibitive_wv', 'custom_dict_oblique_wv',
'custom_dict_cautionary_wv', 'custom_dict_concerned_wv','liwc_risk'])
```

3.3 Reducing safety silence

The above indicated that safety silence is not silent, and that participants who withhold concerns do utter less safety voice speech. The distinction between raising and withholding safety concerns is therefore a matter of degree in terms of uttering speech.

Can therefore interventions increase the extent that participants utter safety voice speech? The hypotheses related to this were worded as below:

- H2a) "Safety silence may therefore be reduced through increasing the salience of hazards (hypothesis 2a)."
- H2b) "Safety silence may be reduced through increasing the extend people feel responsible for the outcomes of hazardous situations (hypothesis 2b)."
- H2c) "Safety silence is reduced through providing encouragements for safety voice (hypothesis 2c)."

The analyses below are prestend as follows:

1. Analyses for direct effects (grouped: H3a-c).
2. Analysis of effect of safety concerns on safety voice (baseline analysis for 3-5).
3. Analysis of effect modification of hazard salience (H2a).
4. Analysis of effect modification of responsibility (H2b).
5. Analysis of effect modification of encouragements (H2c).

The results below that only encouragements have a direct (and favourable) effect on whether participants scored higher on the safety voice dictionaries (analysis 1).

Furthermore, analyses 2-4 below indicate that stronger safety concerns predict more words on the safety voice dictionary (the composite measure, $p = .043$), but only if hazards were salient ($p < .050$) and encouragements provided ($p = .05$). A marginally significant effect suggested clearer responsibilities may also modify this effect ($p = .080$).

This supports effect modification for the encouragemetns and hazard salience manipulations (and indicates this for responsibility) on the relationship between holding stronger concerns and raising/witholding these.

In [ ]:

```
# 1. Analyses for direct effects (grouped H2a-c).
multLinR(dfconcerned, ['condition_awareness','condition_responsibility','condition_discourage'], '
custom_dict_voicecomposite_wv')
```

In [ ]:

```
# 2. Analysis of effect of safety concerns on safety voice
multLinR(df, ['concerned'], 'custom_dict_voicecomposite_wv')
```

In [ ]:

```
# 3. Analysis of effect modification of hazard salience (H2a).
## Hazard salience reduced silence about safety through modifying the effect of safety concern on
speaking-up.
moderationAnalysis(df, 'concerned', 'condition_awareness', 'custom_dict_voicecomposite_wv')
```

In [ ]:

```
# 4. Analysis of effect modification of responsibility (H2b).
## Responsibility tended towards reducing silence about safety through modifying the effect of saf
ety concern on speaking-up.
moderationAnalysis(df, 'concerned', 'condition_responsibility', 'custom_dict_voicecomposite_wv')
```

In [ ]:

```
# 5. Analysis of effect modification of encouragements (H2c).
## Encouragements reduced silence about safety through modifying the effect of safety concern on s
peaking-up
moderationAnalysis(df, 'concerned', 'condition_discourage', 'custom_dict_voicecomposite_wv')
```

Probing effects

The below enables the probing of the effect modification by manipulation. Its is included in the manuscript, but commented out due to the length of the output.

Effect modification analyses were performed for each safety voice dictionary.

1. Effect modification for specific safety voice dictionaries (hazard salience).
2. Effect modification for specific safety voice dictionaries (responsibility).
3. Effect modification for specific safety voice dictionaries (encouragements).

Please note: Due to extensive output the non-significant analyses have been commented-out. These analyses can, of course, be performed through removing the hashtag.

In [ ]:

```
#Frequencies voice stage groups  (including 0)
df['voice_stage'].value_counts()
```

In [ ]:

```
# 1. Effect modification for specific safety voice dictionaries (hazard salience).
# (insignificant and marginal effects are commented out for succinctness but can be performed thro
ugh removing the '#').

#moderationAnalysis(df, 'concerned', 'condition_awareness', 'custom_dict_informative_wv')# Margina
l (only salient)
moderationAnalysis(df, 'concerned', 'condition_awareness', 'custom_dict_inquisitive_wv')# Yes
(only salient)
#moderationAnalysis(df, 'concerned', 'condition_awareness', 'custom_dict_prohibitive_wv')# Margina
l (only salient)
#moderationAnalysis(df, 'concerned', 'condition_awareness', 'custom_dict_cautionary_wv')# No
#moderationAnalysis(df, 'concerned', 'condition_awareness', 'custom_dict_oblique_wv')# No
```

In [ ]:

```
# 2. Effect modification for specific safety voice dictionaries (responsibility).
# (insignificant and marginal effects are commented out for succinctness but can be performed thro
ugh removing the '#').

#moderationAnalysis(df, 'concerned', 'condition_responsibility', 'custom_dict_informative_wv')# No
moderationAnalysis(df, 'concerned', 'condition_responsibility', 'custom_dict_inquisitive_wv')# Yes
(only clear)
#moderationAnalysis(df, 'concerned', 'condition_responsibility', 'custom_dict_prohibitive_wv')# Ma
rginal (only clear)
#moderationAnalysis(df, 'concerned', 'condition_responsibility', 'custom_dict_cautionary_wv')# No
moderationAnalysis(df, 'concerned', 'condition_responsibility', 'custom_dict_oblique_wv')# Yes
(only unclear)
```

In [ ]:

```
# 3. Effect modification for specific safety voice dictionaries (encouragements).
# (insignificant and marginal effects are commented out for succinctness but can be performed thro
ugh removing the '#').

moderationAnalysis(df, 'concerned', 'condition_discourage', 'custom_dict_informative_wv')# Yes
(only discourage)
#moderationAnalysis(df, 'concerned', 'condition_discourage', 'custom_dict_inquisitive_wv')# Margin
al (only discourage)
#moderationAnalysis(df, 'concerned', 'condition_discourage', 'custom_dict_prohibitive_wv')# No
#moderationAnalysis(df, 'concerned', 'condition_discourage', 'custom_dict_cautionary_wv')# No
#moderationAnalysis(df, 'concerned', 'condition_discourage', 'custom_dict_oblique_wv')# No
```

3.4 The effect of time on reducing safety silence

The following analyses were performed to investigate whether safety silence is reduced differently at different time-points of the hazardous scenario (hypothesis 3).

- 1a) Effect of manipulations for the first stage (conceptualisation).
- 1b) Effect of manipulations for the second stage (encounter).
- 1c) Effect of manipulations for the third stage (imminent danger).

Furthermore, the following analyses were performed:

- 2) Differences in safety voice and safety silence across the stages.

Finally, differencesin safety voice dictioary scores were analyses across the stages of the hazard:

- 3a) Differences in safety voice dictionaries for the first stage (conceptualisation).
- 3b) Differences in safety voice dictionaries for the second stage (encounter).
- 3c) Differences in safety voice dictionaries for the third stage (imminent danger).

In summary, these analyses indicated the manipulations reduce silence about safety at different time-points. Safety voice scores were different across these stages of the hazard and participants that spoke-up, spoke differently dependent on the stage of the hazard.

In [ ]:

```
# Data preperation and description.
print('Full dataset:')
df['voice_stage'].value_counts()
print('Dataset without silence and post-hoc:')
dfvoice['voice_stage'].value_counts()
```

In [ ]:

```
# 1a. Effect of manipulations for the first stage (conceptualisation).
outcome = 'conceptvoice'
predictors = ['condition_awareness','condition_responsibility','condition_discourage']

logisticR(df, predictors, outcome, odds_ratio=True)
```

In [ ]:

```
# 1b. Effect of manipulations for the second stage (encounter).
outcome = 'encountervoice'
predictors = ['condition_awareness','condition_responsibility','condition_discourage']

logisticR(df, predictors, outcome, odds_ratio=True)
```

In [ ]:

```
# 1c. Effect of manipulations for the third stage (imminent danger).
outcome = 'hazardvoice'
predictors = ['condition_awareness','condition_responsibility','condition_discourage']

logisticR(dfconcerned, predictors, outcome)
```

In [ ]:

```
# 2. Differences in safety voice and safety silence across the stages.
anova(dfvoice, 'voice_stage', 'custom_dict_voicecomposite_wv')
manova(dfvoice, ['voice_stage'], ['custom_dict_concerned_wv','custom_dict_disfluency_wv',
'custom_dict_informative_wv', 'custom_dict_inquisitive_wv', 'custom_dict_prohibitive_wv',
'custom_dict_cautionary_wv', 'custom_dict_oblique_wv'])
```

In [ ]:

```
# 3a) Differences in safety voice dictionaries for the first stage (conceptualisation).
manova(dfvoice, ['conceptvoice'], ['custom_dict_concerned_wv','custom_dict_disfluency_wv',
'custom_dict_prohibitive_wv','custom_dict_informative_wv', 'custom_dict_inquisitive_wv',
'custom_dict_cautionary_wv', 'custom_dict_oblique_wv'])
```

In [ ]:

```
# 3b) Differences in safety voice dictionaries for the second stage (encounter).
manova(dfvoice, ['encountervoice'], ['custom_dict_concerned_wv','custom_dict_disfluency_wv',
'custom_dict_informative_wv', 'custom_dict_inquisitive_wv', 'custom_dict_prohibitive_wv',
'custom_dict_cautionary_wv', 'custom_dict_oblique_wv'])
```

In [ ]:

```
# 3c) Differences in safety voice dictionaries for the third stage (imminent danger).
manova(dfvoice, ['hazardvoice'], ['custom_dict_concerned_wv','custom_dict_disfluency_wv',
'custom_dict_informative_wv', 'custom_dict_inquisitive_wv', 'custom_dict_prohibitive_wv',
'custom_dict_cautionary_wv', 'custom_dict_oblique_wv'])
```