Spring 2024

Passenger Traits That Predict the Bystander Intervention Model Steps During an Inflight Assault on a Flight Attendant

Michael Pettit
Embry-Riddle Aeronautical University, petti14c@my.erau.edu

Passenger Traits That Predict the Bystander Intervention Model Steps During an Inflight Assault on a Flight Attendant

Michael Robert Pettit

Dissertation Submitted to the College of Aviation in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Aviation

Embry-Riddle Aeronautical University

Daytona Beach, Florida

April 2024

Passenger Traits That Predict the Bystander Intervention Model Steps During an Inflight Assault on a Flight Attendant

By

Michael Robert Pettit

This dissertation was prepared under the direction of the candidate's Dissertation Committee Chair, Dr. Scott R. Winter , and has been approved by the members of the dissertation committee. It was submitted to the College of Aviation and was accepted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Aviation.

Scott R. Winter Digitally signed by Scott R. Winter Date: 2024.04.29 14:55:15 -04'00'

Scott R. Winter, Ph.D.
Committee Chair

Francis H Ayers Jr. Digitally signed by Francis H Ayers Jr. Date: 2024.04.29 14:59:43 -04'00'

Frank H. Ayers, Jr., Ed.D.
Committee Member

Steven Hampton Digitally signed by Steven Hampton Date: 2024.04.30 11:53:27 -04'00'

Steven Hampton, Ed.D.
Associate Dean, School of Graduate Studies, College of Aviation

Joseph Keebler Digitally signed by Joseph Keebler Date: 2024.04.30 08:50:06 -04'00'

Joseph R. Keebler, Ph.D.
Committee Member

Alan J. Stolzer Digitally signed by Alan J. Stolzer Date: 2024.04.30 14:45:26 -04'00'

Alan J. Stolzer, Ph.D.
Dean, College of Aviation

Dothang Truong Digitally signed by Dothang Truong Date: 2024.04.29 15:32:21 -04'00'

Dothang Truong, Ph.D.
Committee Member

EDWIN ODISHO Digitally signed by EDWIN ODISHO Date: 2024.04.29 16:39:29 -04'00'

Edwin Odisho, Ph.D.
Committee Member (External)

April 29, 2024

Signature Page Date

Abstract

Researcher: Michael Robert Pettit

Title: Passenger Traits That Predict the Bystander Intervention Model Steps During an Inflight Assault on a Flight Attendant

Institution: Embry-Riddle Aeronautical University

Degree: Doctor of Philosophy in Aviation

Year: 2024

Every major U.S. airline has experienced a violent inflight assault on one of their flight attendants. The problem is becoming more widespread, but industry leaders have few coping strategies. The first step toward addressing the issue is understanding the various aspects of the event, including the actions of other passengers (bystanders) who may witness the assault. There is ample literature on bystander reactions to similar events such as bullying, medical emergencies, injuries, and violent attacks, but none dedicated to a flight attendant assault or passenger reactions to it.

At the core of bystander action is the bystander effect, which postulates the inhibiting effect that others have on a person's behavior. Overcoming the inhibition involves a five-step cognitive process. Researchers have modeled and positively applied the process, known as the bystander intervention model, to various helping scenarios (bullying, sexual assault, organ donating, and environmental conservatism) but have thus far left flight attendant assaults unresearched.

The research used partial-least-squares structural equation modeling to expand the understanding of passenger-bystanders and the bystander intervention model by evaluating its applicability to a flight attendant assault and identifying predictors of its steps. Data was collected from adult American air travelers to assess the relevance and

strength of the relationships between model steps, their predictors, and a passenger's likelihood to intervene during an assault. Most outer models showed moderately strong reliability and validity, although step three of the bystander intervention model was removed for having low discriminant validity. The remaining steps showed significant ability to predict their following steps, and only the expectation of a positive outcome was a nonsignificant predictor. However, the resulting disjointedness of the model constrains assigning its relevance to a flight attendant assault. Also included was an evaluation of differences between males and females for those outer models found to be at least partially measurement invariant. This evaluation showed mixed results, attesting to the complexities of the gender variable in bystander studies. The findings broaden the research application of the bystander intervention model and allow practitioners to develop mitigation strategies.

Keywords: flight attendant, air rage, bystander, bystander intervention model, inflight assault, airline, passenger, structural equation modeling, PLS-SEM

Table of Contents

Page

List of Tables

List of Figures

Chapter I: Introduction

The first chapter will provide background information on a recently-highlighted airline industry problem. The chapter explains how the research is related to addressing the issue and how it can be helpful to stakeholders. The two research questions are stated, as are the associated hypotheses. The chapter also briefly defines the 12 variables and closes by explaining the delimitations, limitations, and assumptions.

Background

In January 2016, a woman flying on a United Airlines flight to Chicago physically attacked a female flight attendant and caused injuries to her face (CBS New York, 2016). One man, seated nearby, jumped up to grab the attacker's arms, while two others tried to hold her legs. In October 2020, a similar event occurred on a Delta Air Lines flight in which a woman attacked a female flight attendant after a dispute over seatbelts and face coverings (Stanwood, 2020). Passengers seated nearby looked on, with some even filming the assault. Only one man eventually intervened, placing himself between the two women. In 2021, American Airlines experienced at least two incidents where women attacked female flight attendants. One incident potentially involved a mental condition (Associated Press, 2021), while the other was partially triggered by the flight attendant failing to collect the woman's garbage (WABC, 2021). In the latter incident, the attacker followed the flight attendant to the aircraft galley and began punching her, causing several injuries to the flight attendant's arms, legs, face, and neck (WABC, 2021). Only one passenger—an off-duty police officer—came to the victim's aid, restraining the attacker until the plane landed. In another 2021 incident, a woman flying on Southwest Airlines violently assaulted a female flight attendant after being told to return to her seat

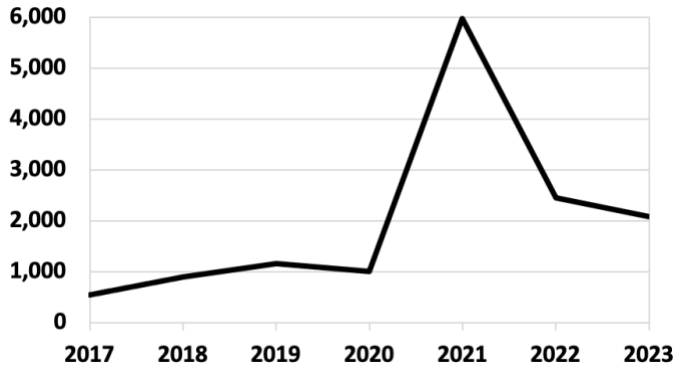
(CBSLA Staff, 2021). An eyewitness to the assault later told reporters, “I just want to know where are the people that should’ve in the back [*sic*] stand up and try to protect that flight attendant?” (CBSLA Staff, 2021, p. 1).

Airlines reported over 5,980 incidents of unruly passengers in 2021—a record-high number (FAA, 2021a). The amount increased by over 500% compared to previous years (Fitz-Gibbon, 2021), and dealing with an unruly passenger has become a top concern for flight attendants (Bell, 2022). Although incidents decreased in 2022 (to 2,359), the problem persists (FAA, n.d.), and the severity of incidents is increasing (ICAO, 2019). The issue creates a dilemma for airline managers, who are often hesitant to report such incidents for fear of their brand becoming associated with passenger violence (Dahlberg, 2016; Luckey, 2000).

While the dimensions of unruly behavior are broad, the most serious (short of a hijacking) include violent assaults on flight attendants. Prohibitions against assaulting or threatening a flight attendant are codified in both U.S. law (see Transportation, 2015) and federal regulation (see Prohibition on Interference with Crew Members, 2023). Possibly due to the historical rarity of such events, the Federal Aviation Administration (FAA) does not publish historical data on the number of flight attendant assaults. However, they provide the number of unruly passenger events reported (see Figure 1).

Figure 1

Number of Unruly Passenger Reports



Note. The FAA does not provide data on flight attendant assaults or severity of unruly passenger incidents, only how many are reported. Adapted from "Unruly Passenger Statistics" by Federal Aviation Administration, n.d., *Dangerous Behavior Doesn't Fly* (https://www.faa.gov/unruly).

The very nature of a passenger airliner means that in all cases of flight attendant assaults, there are witnesses—or bystanders—to the event. In fact, being an aircraft passenger makes being a bystander unavoidable, if not in terms of witnessing the event, then at least in terms of proximity and shared environment. As seen via media reporting of the incidents, passenger-bystanders often avoid intervening and sometimes even video the event, which may imply approval. What is also evident from reporting is that some passengers occasionally intervene, although it is unclear what compels them to do so. As in many victimization cases, people question why more passenger-bystanders do not intervene (see CBSLA Staff, 2021).

At the operational level, flight attendant assaults can cause flight diversions, disrupted travel, significant delays, and increased passenger costs (McLinton et al., 2020). At the personal level, an assault can cause severe injuries and have long-term

emotional impacts, not only for the victim (Akgeyik, 2011; Gale et al., 2018; Hu et al., 2017; Williams, 2000) but for other passengers as well (Pierson et al., 2007; Rhoden et al., 2008). The proliferation of flight attendant assault stories in the news may also have a marketing impact, as it creates an unfavorable impression of air travel—something airlines attempt to portray as a pleasant experience (Bor & Van Gerwen, 2003; Budd, 2011; Morgan & Nickson, 2001). Despite those attempts, airlines and their employees discreetly prepare for passenger conflict (Pawlowski, 2013; Spells et al., 2021).

During boarding, flight attendants mentally identify passengers they think might help restrain an unruly person (Pawlowski, 2013). They commonly base their choice on perceived skills and personality (i.e., military personnel, off-duty law enforcement officers) or physical characteristics (size and strength). However, research into bystander intervention reveals that the influence of such characteristics on helping is often moderated by other characteristics of the victim, perpetrator, bystander, situation, and environment (Brewster & Tucker, 2016). While the selection criteria are certainly noteworthy, what is pertinent to this research is that flight attendants clearly regard passenger-bystanders as a helping resource (see Bailey, 2022; see Spells et al., 2021; see Wallace & Muntean, 2021; see Zhao, 2021), even though anecdotal evidence shows that resource to be unreliable.

Statement of the Problem

Federal law and aviation regulations specifically address and prohibit the assault of a flight attendant (see Transportation, 2015; see Prohibition on Interference with Crew Members, 2023). However, such events have increased, garnering national media attention. What is evident from the news stories is that despite dozens of bystanders, the

flight attendant victim is often left to defend themself unaided. Although research has covered a wide range of victimization types and scenarios, none have included the inflight environment, the social dynamics of a commercial aircraft, or analysis of passenger intervention. The omission creates a gap in understanding these events and the behavior of those who witness them.

A literature review highlights a gap in the body of knowledge covering the social dynamics in an airline passenger cabin. Existing research primarily emphasizes the business aspects such as service expectations, customer loyalty, or product value. However, the characteristics of an airline passenger cabin are unique, as are the interpersonal requirements of its occupants. Because of the environment, behavioral expectations, and social roles, an assault on a flight attendant presents a situation different from any other violent event. Research into the event as a social phenomenon, however, is nonexistent.

Purpose

Due to a rise in unruly passenger behavior, the FAA and the commercial airline industry have begun to pay closer attention to the flight attendant assault phenomenon (see Dickson, 2021). However, the research community has left the phenomenon and the behavior of its bystanders unaddressed. The purpose of this research is to expand the understanding of the bystander intervention model (BIM) (see Latané & Darley, 1970) by evaluating its applicability to a hypothetical scenario and identifying factors found to be relevant in related research. To do this, the researcher used a quantitative, nonexperimental design to assess the influence of passengers' characteristics on their intention to intervene on behalf of a flight attendant during a physical assault. The results

further increase government and private stakeholders' understanding of the phenomenon and passenger-bystander behavior. Such greater understanding can lead to more effective incident-reduction strategies.

The research collected participants' self-assessment of their intrinsic characteristics and their intention to accomplish the steps of the BIM when faced with a written, scenario-based description of a flight attendant being assaulted. These factors were used in a structural equation model to evaluate their influences. In addition, the researcher conducted a between-groups (male and female) analysis to determine whether significant differences exist.

Although federal and state criminal and civil laws include various definitions of *assault*, they are highly legalistic and multi-faceted, including such aspects as intent, imminence, and reasonable perception of threat and harm (see GA Code, 2022; see FL Statute, 2023). The scope of the research was not whether participants were aware of the legal facets of the event or even whether the event rose to the level of illegal behavior. Instead, the purpose was to evaluate participants' responses to it. Similarly, government directives responding to the COVID-19 outbreak (see Transportation Security Administration [TSA], 2021) likely influenced many of the assaults during the research period. Flight attendants were tasked with enforcing often unpopular restrictions, which may have triggered many assaults. However, the scope of the research lay outside any specific impetus to the assault and only focused on the passenger's reaction to it. The methodology included presenting a scenario in which participants, as passengers, witness an assault but are unaware of what led to its occurrence. In essence, it was the assault itself that commanded the passenger's attention, not what caused it.

Significance of the Study

Practitioners need to understand both the conceptual and practical aspects of a problem before adopting effective solutions. This research contributes to the conceptual understanding of the inflight assault phenomenon by modeling the behavior of those who would witness the event. It explains how recognition, responsibility, knowledge, and decision-making relate to passenger-bystander behavior and quantifies the relevance of intrinsic characteristics that may influence a person's willingness to intervene. In a broader context, doing so furthers the understanding of violent assaults and the bystander effect and expands the theoretical boundaries of the BIM. The increased understanding of violent assaults and the bystander effect can be the foundation for stakeholders to develop mitigation strategies specific to flight attendant assaults.

Using the results, airlines may be able to prepare flight attendants more thoroughly regarding expectations of assistance. They may also be able to provide flight attendants with actions they can take to increase prosocial passenger-bystander behavior. For example, measures that reduce the ambiguity of an event and more clearly convey that help is needed may facilitate more intervention. Regarding preselecting passengers who might assist, airlines may use the results to design better flight attendant training programs. Although flight attendants cannot identify a passenger's latent qualities, they can be informed of the unreliability of traditional selection criteria.

Similarly, government stakeholders might also use the results to design public-awareness campaigns (see European Union Aviation Safety Association [EASA], n.d.; see McLinton et al., 2020) that capitalize on factors that influence intervention. For example, the FAA publishes many informational airport signs targeting potential assailants (see

FAA, 2021b), but none targeting potential helpers. Just as anti-human-trafficking campaigns attempt to increase traveler awareness and involvement, so too may bystander campaigns.

Research Questions

The study answered the following research questions:

RQ_1.

Which steps of the bystander intervention model represent the relationships between existing theoretical factors and a passenger's willingness to intervene during an inflight assault on a flight attendant?

RQ_2.

Which factors significantly predict a passenger's willingness to intervene during an inflight assault on a flight attendant?

Hypotheses

Originally, 10 hypotheses postulated positive relationships between variables in the context of a flight attendant assault. Due to alterations of the research model (based on discriminant validity analysis), two hypotheses were added. The original 10 are presented here with an explanation of the additions provided in Chapters IV and V. For each relationship, there is a subset hypothesizing that the relationship differs significantly between males and females. The study assumes null hypotheses, but they are omitted for brevity. Research variables are described in a subsequent section.

H_1

Recognizing the need for intervention significantly positively influences taking responsibility.

$H_{1\text{-}1}$

There is a significant difference between males and females in the relationship between recognizing the need for intervention and taking responsibility.

H_2

Taking responsibility significantly positively influences knowing what to do.

$H_{2\text{-}1}$

There is a significant difference between males and females in the relationship between taking responsibility and knowing what to do.

H_3

Knowing what to do significantly positively influences deciding to act.

$H_{3\text{-}1}$

There is a significant difference between males and females in the relationship between knowing what to do and deciding to act.

H_4

Nonintervention cost significantly positively influences recognizing the need for intervention.

$H_{4\text{-}1}$

There is a significant difference between males and females in the relationship between nonintervention cost and recognizing the need for intervention.

H_5

Out-group social bias significantly positively influences taking responsibility.

$H_{5\text{-}1}$

There is a significant difference between males and females in the relationship between out-group social bias and taking responsibility.

H_{6}

Intervention skills significantly positively influences knowing what to do.

$H_{6\text{-}1}$

There is a significant difference between males and females in the relationship between intervention skills and knowing what to do.

H_{7}

Self-efficacy significantly positively influences knowing what to do.

$H_{7\text{-}1}$

There is a significant difference between males and females in the relationship between self-efficacy and knowing what to do.

H_{8}

Perceived social influence significantly positively influences taking responsibility.

$H_{8\text{-}1}$

There is a significant difference between males and females in the relationship between social influence and taking responsibility.

H_{9}

Perceived social influence significantly positively influences deciding to act.

$H_{9\text{-}1}$

There is a significant difference between males and females in the relationship between social influence and deciding to act.

H_{10}

Expectation of positive outcome significantly positively influences deciding to act.

$H_{10\text{-}1}$

There is a significant difference between males and females in the relationship between expectation of positive outcome and deciding to act.

Description of Variables

Table 1 lists the variables, their usage, and their description. Many of the descriptions were adapted from similar research to be non-scenario specific. The variables in the structural equation model form are presented in Chapter II.

Table 1

Research Variables and Descriptions

Variable	Operational Description
Exogenous	
Nonintervention cost[a]	Potential negative impacts to a bystander due to an event's continuance.
Out-group social bias	Favorable views extended toward members of a group to which the observer does not belong.
Intervention skills	Specialized prosocial training.
Expectation of positive outcome[b]	The expectation that the result of intervention will be positive.
Situation outcome[b]	The expectation that intervention will result in conflict diffusion.
Victim outcome[b]	The expectation that intervention will result in reduced feelings of victimization.
Self-efficacy	An individual's belief in their ability to perform effectively.
Perceived social influence	A person's perception of how friends and parents expect them to act.
Endogenous	
Recognizing the need for intervention	The interpretation of an event as one that will not be resolved without assistance.
Taking responsibility	A bystander's acceptance of personal responsibility to intervene in an event.
Knowing what to do	Identifying an intervention strategy that the bystander believes will be effective.
Deciding to act	The intention to implement an active intervention strategy.

Note. [a] Related research includes this variable as a victim-centered one—the cost to the victim if bystanders do not intervene (death, injury, humiliation). Other research includes a nonintervention cost to the bystander, but it is not measured. Instead, such research assumes the cost is present due to the nature of the event, such as a building fire.

[b] *Situation outcome* and *victim outcome* are lower-order constructs. *Expectation of positive outcome* is the higher-order construct.

Delimitations

Participants were not physically immersed in the inflight event. The study was delimited to an attitudinal assessment of participant responses and their perceptions of their actions when presented with the flight attendant assault scenario. To keep the written scenario to a reasonable length and avoid participant inattention, the scenario omitted describing ancillary details of the environment and its actors (flight attendant, attacker, and other bystanders). Although doing so left some variables imagined and uncontrolled, it kept the scenario from being overly specific and exclusive. Also, the victim and assailant details (female victim, female assailant) are only one of four possible gender combinations. Whether or not the combination comprises a proportionate of attacks is unknown, as there is no research or available data on the topic. Instead, researchers are left to review news reports to understand the details of the events. Such a review revealed that the combination of female attacker and female victim is a common one (see Associated Press, 2021; see CBS New York, 2016; see CBSLA Staff, 2021; see Stanwood, 2020; see WABC, 2021). As such, it was used as the research scenario, even though it limits the generalization of findings to the other three assault combinations. However, since the topic is completely absent in the literature, selecting a familiar

combination establishes an initial scope for follow-up researchers to expand. Such researchers can use results as a baseline to conduct research using assailant gender and flight attendant gender as variants.

Limitations

The constructs included in this research are relatively numerous, but the list is not all-inclusive. The researcher selected the 12 constructs based on their repeated emergence in parallel bystander research. While other variables may also influence a passenger's willingness to intervene, specifying an excessive number may reduce the model's explanatory power. As in any bystander research, the inability to include an exhaustive list of relevant factors limits the power to draw definitive conclusions about or predict, with complete accuracy, a participant's behavior. Also, some participants may have witnessed an actual flight attendant assault and defaulted to describing their previous actions instead of the scenario-based assessment. However, the number of such participants is expected to be extremely small.

Assumptions

A participant's responses were assumed to be truthful to the extent that they could accurately predict their behavior. Although collection occurred without in-person contact, the answers were assumed to be honest and intentional. In addition, the researcher assumed each participant submitted their data without collaboration and in an attentive manner or that the impact of failing to do so was negligible. Lastly, although the data was collected via a third-party interface (an online survey platform), the researcher assumed the raw data to be uncorrupted.

Summary

Chapter I introduced real-world events, highlighting how the topic has become one in need of attention. The background information was summarized into a problem statement, and the research purpose was explained. There is both a social science and practical significance to the research, so both were explained. The two research questions and all hypotheses were listed, as was a description of the variables. The chapter closed with a candid assessment of delimitations, limitations, and assumptions.

Definition of Terms

Benevolent sexism	Sexism beneficial to the opposite sex, such as a male feeling a sense of duty to protect and care for women (Leone et al., 2020).
Chat room	See social networking website.
Closed-circuit television	A television system in which signals are not distributed publicly but are instead used for surveillance and security (TechTarget, 2021).
Cyber-bullying	Bullying over electronic or digital medium as opposed to in-person (DeSmet et al., 2016).
General self-efficacy	An individual's belief in their ability to perform effectively in general circumstances (Bandura, 1997).
Green living	Environmental actions taken for the good of the public domain (waste management, recycling, consumption reduction) (Anker & Feeley, 2011).

Hierarchical component model	A measurement model in which two or more lower-order components comprise the dimensions of a single, multidimensional higher-order construct (Hair et al., 2014).
In-group	A social group in which a person is a member (Hewstone et al., 2002).
Out-group	A social group in which a person is not a member (Hewstone et al., 2002).
Peer pressure	Pressure from peers to act or refrain from acting regardless of personal desires (adapted from Clasen & Brown, 1985).
Prosocial	Interaction between a benefactor and recipient that is characterized by society as being beneficial (Dovidio et al., 2017).
Social networking website	An internet service that allows users to connect with other users, post content, and comment either publicly or privately (Orlando, 2020).
Social status	A ranking of individuals or groups within society (either in isolation or in relation to oneself) based on traits, assets, and roles (adapted from Weiss & Fershtman, 1998).
Specific self-efficacy	An individual's belief in their ability to perform a specific task effectively (DeSmet et al., 2016).

List of Acronyms

AVE	Average variance extracted
BIM	Bystander intervention model
CB	Covariance-based
CCTV	Closed circuit television
CFA	Confirmatory factor analysis
EFA	Exploratory factor analysis
HCM	Hierarchical component model
HOC	Higher-order construct
HTMT	Heterotrait-monotrait ratio
I-CVI	Item content validity index
IRB	Institutional review board
LM	Linear regression model benchmark
LOC	Lower-order constructs
MAE	Mean absolute error
MGA	Multigroup analysis
MICOM	Measurement invariance of composite models
MTurk	Amazon Mechanical Turk
PLS	Partial least squares
RMSE	Root mean square error
SEM	Structural equation modeling
SME	Subject matter experts
VIF	Variance inflation factor

Chapter II: Review of the Relevant Literature

Chapter II begins by briefly reviewing the inflight violence phenomenon. It then examines bystander behavior and its underlying theoretical foundation, distinctive elements, and recent variations to the theoretical basis. The principal theoretical model and the constructs included in the research model are explained. Existing findings related to gender and the research gaps are presented. The chapter closes with an explanation of the research hypotheses.

Violence in the Air

Aggressive behavior occurring on an aircraft is unique. Collectively known as *air rage* (see McLinton et al., 2020; see Rhoden et al., 2008; see Tsang et al., 2018; see Vredenburgh et al., 2015), such behavior occasionally escalates to the point of physical violence. Although the term is routinely used in media (Rhoden et al., 2008), airline and government agencies prefer more sterile terms such as *unruly* or *disruptive passenger behavior* (Dahlberg, 2016). Even though the trigger to such behavior is usually attributed to an onboard event, it may actually be extended in its development, occurring only at the end of an enduring series of stressors unique to air travel. These stressors may amplify psychological tendencies and cause a person to lash out in ways most others, even themselves, would typically not.

Situational Stress

An interpersonal violent act usually comes at the end of several micro-agitating but compounding events (see Rhoden et al., 2008). For inflight incidents, what results in physical violence is typically the manifestation of multiple stressful events in a long travel process (Nelms, 1998). Even before arriving at the departure gate, travelers often

suffer from sleep deprivation, traffic frustration, and airport security aggravation (Genç & Dural, 2009; Richards et al., 2016). After reaching their gate, travelers are commonly met with additional stressors, such as oversold seats, long lines at the check-in counter, and delays (DeCelles et al., 2019). The stress can become further elevated after boarding, as passengers are then confronted with crowding, uncomfortable temperatures, lack of courtesy from other passengers (see Vredenburgh et al., 2015), crying babies (DeCelles et al., 2019), small seat size, limited leg room (McLinton et al., 2020), and an encroachment on their interpersonal space (DeCelles & Norton, 2016; McLinton et al., 2020; Vredenburgh et al., 2015; Whitley & Gross, 2019). The combined stressors can cause travelers to feel like they have lost control over their surroundings, comfort, privacy, and autonomy. The result may be a passenger becoming extremely agitated—far exceeding normal levels (see Vredenburgh et al., 2015)—and poised for conflict with the next person, any person, who further limits their control. At this point, the flight attendant is unknowingly forced into a precarious predicament since much of their job involves control.

From the moment a passenger enters the cabin, flight attendants have control over their seating, freedom of movement, entertainment, and access to possessions. A flight attendant insisting a passenger stow belongings, cease phone communication, or use a face-covering may unknowingly impart the final trigger to a violent response. Flight attendants may attribute the conflict to the interaction (see McLinton et al., 2020), but it likely begins several hours earlier and may even include an unseen psychological component.

Psychological Stress

Although research has identified travel stress as leading to passenger violence, it may not be the exclusive cause. After all, millions of travelers each day experience the same stress, but very few turn to violence. Triggering an aggressive response may require adding some psychological change as well. It may not solely be the travel stress that incites violence, but instead, its mixing with an emotional state such as claustrophobia, flight anxiety, depression (Baranishyn et al., 2010; Bor, 2007; Budd, 2011; Oakes & Bor, 2010), agoraphobia (Tsang et al., 2018), severe discomfort (such as a headache) (Genç & Dural, 2009; Richards et al., 2016), or general resistance to authority. This final tendency makes the flight attendant a particularly viable target, as they clearly represent *the authority* in the aircraft cabin. Unfortunately, such various hidden conditions make identifying potential aggressors difficult, especially for a flight attendant who is only briefly interacting with each person. These conditions have been scarcely included in passenger research.

Researchers have been slow to include psychological variables in their methods due to the sensitivity of collecting data on passengers' emotional states (Tsang et al., 2018). Instead, research has concentrated on expanding the list of influential overt characteristics of travel and the physical and social environment in the aircraft cabin (see Tsang et al., 2018). The relatively narrow focus leaves a significant gap in understanding how psychological tendencies, exacerbated by travel stress, might cause a passenger to become a violent attacker.

Airline Passenger-Attackers

Characterizing those who commit violent acts airborne is difficult, as no stereotypical offender exists (McLinton et al., 2020). Researchers have identified some principal traits such as being demanding and intolerant (Akgeyik, 2011; Bor, 2003; Bor et al., 2001; Salinger et al., 1985; Smart & Mann, 2003), anxious and anger-prone (Bricker, 2005; McIntosh et al., 1998; Menon & Dubé, 2004; Menon & Dubé, 2007), or anti-social (Meldrum, 2016). However, none of the above findings are remarkable, as they do not distinguish passenger offenders from those in other contexts. The most commonly found characteristic in incidents is not biological or psychological but behavioral—the use of alcohol by the offender (Anglin et al., 2003; Barron, 2002; Berkley & Ala, 2001; Bor et al., 2001; Cook, 1997; Girasek & Olsen, 2009; Smart & Mann, 2003). Although the finding is not surprising, it is notable because while most travel stressors (airport traffic congestion, crowding, security requirements) and psychological states are difficult or impossible to control, the availability of alcohol is wholly within the purview of airport authorities and airlines. Such a realization, however, creates an added dilemma for managers, as it ties potential violence-reduction to profit. Airline and airport managers must decide how much revenue loss is acceptable to prevent flight attendants from being victimized by intoxicated passengers.

Flight Attendant Victims

A flight attendant being victimized creates an uncommon inversion of social roles. Before an assault, a flight attendant holds a position of higher control and social power. During the initial stages of an assault, however, power and control transfers to the attacker. The shift makes the flight attendant's response much more complicated than

other types of victims. The flight attendant victim must not only defend themselves (physical response) but also regain control and power (social response). The bidimensional response is vital not only for the flight attendant's safety but for the safety of other passengers as well since they may perceive the flight attendant as having lost control and authority. Having to regain that authority makes achieving a successful outcome much more critical than that required by other victims in similar scenarios. However, practitioners have thus far been primarily concerned with the success of the physical response.

To assist in defending against passenger-attackers, federal air marshals have begun to offer free self-defense training for flight attendants (Aratani, 2016). Unfortunately, such a step belies stakeholder acquiescence that exposure to violence has become an expected hazard for cabin crews, particularly females (see Ballard et al., 2006; see Pontell et al., 1983). Whereas male flight attendants are very rarely attacked by female passengers, female flight attendants do not benefit from the opposite-gender exclusion (Akgeyik, 2011). In addition, while both groups experience a similar degradation in work performance after an assault, the enduring adverse physiological impact on female victims is more severe (Williams, 2000). The findings reveal that in terms of negative impacts, female flight attendants are clearly the more victimized group. Females are also less apt to initially respond assertively, which may mean they inadvertently allow the incident to escalate to the point where conflict-reduction strategies are ineffective and physical restraint is needed (see Braithwaite, 2001), sometimes even necessitating help from other passengers.

Other Passengers

When passengers witness violent outbursts, they usually attribute the behavior to a lack of self-control, a decline in personal responsibility, and an increased sense of entitlement (Small & Harris, 2018). If such opinions persist, passengers who witness violent outbursts will be less likely to attribute it to situational or psychological stress and more likely to assign blame directly to the attacker. It may be that other passengers—those who are better able to tolerate travel stress—are wholly unsympathetic to those who cannot. Most travelers likely accept situational and psychological stress as an inherent part of air travel and consent that being able to cope is a prerequisite to becoming an airline passenger.

Airline Passenger-Bystanders

Passenger-bystanders are unique. The seating density and confinement create a much more intimate bystander experience than that felt by traditional onlookers. Even though assailants commit many violent assaults in the presence of witnesses (Hart & Miethe, 2008), they commit very few within an arm's reach of several dozen bystanders. Even in the most crowded bar or stadium, a bystander can push their way clear of the altercation. In an aircraft, even if only partially filled, passenger-bystanders are trapped in the event. These characteristics make the experience of the passenger-bystander unique from their counterparts in ground-based events.

Theoretical Foundation – The Bystander Effect

The inhibiting effect that the presence of others has on a person's inclination to help is known as the *bystander effect* (Dovidio et al., 2017). Research into the phenomenon began in the mid-60s after the high-profile Kitty Genovese case in which a

young woman was stalked, attacked twice, and stabbed to death over a half-hour outside her New York apartment complex (Brewster & Tucker, 2016). The following day, news outlets reported that up to 38 witnesses saw the attack but failed to call the police or assist (Thomas, 2018). Although these initial reports contained a significant number of errors (Thomas, 2018), what is true is that only one person came directly to Genovese's aid, albeit 30 minutes after the attack began (Ruhl, 2021). From that event, research into the phenomenon of the passive bystander ensued.

Many attributed the actions of the Genovese onlookers to the simple moral decline of urban society (Gallo, 2015). Research suggests that although 65% of violent events occur in the presence of witnesses (Hart & Miethe, 2008; see Planty, 2002), up to 80% of those witnesses will *not* come to the aid of the victim (Nickerson et al., 2014). Early research consistently found that bystanders were less likely to intervene if others were present (Latané & Darley, 1968, 1969, 1970; Latané & Nida, 1981; Latané & Rodin, 1969; Ross, 1971). The inhibition to act increases sharply with the number of onlookers, as the percent of those willing to help has been shown to drop from 85% when alone to 31% when four others were present (Darley & Latané, 1968). These early findings were undoubtedly consistent but misleading since they made little differentiation between events. With so little understanding of bystander behavior, early researchers tended to overgeneralize their findings. The broadest explanation for such findings was a diffusion of responsibility (Darley & Latané, 1968; Latané & Nida, 1981; Morgan, 1978; see Wallach et al., 1964).

Mathematically, if responsibility is equally distributed amongst onlookers, then a greater number of onlookers results in less responsibility felt by each individual (Latané

& Darley, 1970). The level of responsibility attributed to the collective *others* far outweighs the level felt by any individual (Darley & Latané, 1968). The total weight of this responsibility resting on a single bystander makes them more likely to act (Chekroun & Brauer, 2002; Latané & Darley, 1970; Tilker, 1970) and do so more quickly (Darley & Latané, 1968). The explanation theorizes that responsibility exists as an element with a fixed quantity that is not altered in various events but only dispersed. The consistency of these findings led to a search for the boundaries of the bystander effect.

As researchers investigated the extent of the phenomenon, they discovered that the inhibition to help was not limited to violent attacks (Schwartz & Gottlieb, 1976). Instead, the same inhibition manifests itself in a myriad of severe but non-violent scenarios such as an injury (Harris & Robinson, 1973; Latané & Darley, 1968; Latané & Rodin, 1969; Shotland & Heinold, 1985), illness (Darley & Latané, 1968; Harris & Robinson, 1973; Piliavin et al., 1969), stranded motorist (Hurley & Allen, 1974), and theft (Howard & Crano, 1974). Research even confirmed its existence in rather mundane events such as picking up a dropped object (Latané & Dabbs, 1975), offering directions (Allen, 1968; Harada, 1985), or answering a door (Levy et al., 1972; Morgan, 1978). These repeated findings likely solidified the belief in a specific number of bystanders reaching a *shared responsibility* threshold where no more helping would occur. Although not explicitly stated, the focus of early research on varying the type of event suggests an inclination toward this belief. More modern research has broadened the scenarios to include child-abuse (Christy & Voigt, 1994), adolescent bullying (DeSmet et al., 2016; Dillon & Bushman, 2015; Gini et al., 2008; Hawkins et al., 2001), charitable giving (Garcia et al., 2002; Nihan & Gleibs, 2021; Wiesenthal et al., 1983), cyber-bullying

(Bastiaenses et al., 2014; DeSmet, 2016; Obermaier et al., 2014), vandalism (Chekroun & Brauer, 2002), and computer technical-assistance (Markey, 2000). The existence of the phenomenon in the context of a flight attendant assault or any incident onboard an aircraft, however, is unresearched. Instead, researchers have focused on identifying variables that might attenuate or enhance the effect.

Characteristics of the Actors

Since the event which spurned the conceptualization of the bystander effect was a violent human event, researchers often look for the influence of human characteristics on bystander behavior. Specific features such as the height and weight of the attacker and bystander correlate with intervention, with effects often being significant at both the first and higher orders (Huston et al., 1981; Laner et al., 2001). However, results have been diametrically opposite regarding attacker qualities such as physique, aggressiveness, and fierceness. Impacts of these qualities vary between positive, negative, and no effect on bystander intervention (see Allen, 1968; see Austin, 1979; see Fischer et al., 2006). Despite the varied findings, these results, in aggregate, reveal that a bystander's response may not be pre-determined based on some quantifiable level of responsibility. Instead, bystander behavior (both active and passive) appears driven by a host of bystander-centric influences. Even ordinal characteristics of the assailant (height, weight, aggressiveness) are only genuinely definable within the perception of the bystander. Instead of variables such as height, weight, and aggressiveness, more useful descriptions may be taller/shorter, heavier/lighter, and more aggressive/less aggressive. Researchers can even operationalize victim characteristics relative to bystander perception.

Piliavin et al. (1969, 1975) were among the first to go beyond observable attacker characteristics by examining the perceived deservedness of victims. Their field experiment of a person collapsing on a subway found bystanders helped a sober person at twice the rate of an intoxicated one. Bystanders also rendered help more quickly to the sober person (Piliavin et al., 1969). These results are novel in illustrating that if responsibility is, in fact, a fixed quantity with a definite diffusion threshold, the type of event does not solely mark its value. More recent research supported the viability of the deservedness variable by finding that bystanders are more likely to offer help if they perceive the victim as having no control over their predicament (Greitemeyer et al., 2006; Greitemeyer & Rudolph, 2003; Meyer & Mulherin, 1980; Weiner, 1980; see Piliavin et al., 1969).

Before deciding to respond, bystanders may subconsciously evaluate whether a victim is even worthy of being helped (Loewenstein & Small, 2007). Such an evaluation often exists in sexual-assault cases where bystanders have been slow to respond due to the victim's level of intoxication or provocative behavior (Burn, 2009). Genovese was attacked at 3:15 am. Perhaps that fact caused her neighbors to develop false assumptions about her conduct, influencing them to remain passive. The explanation is conjecture, as such evidence was never collected or investigated. A person's social desirability may further temper their level of worthiness.

Victims with a visible deformity are less likely to be helped (Piliavin et al., 1975), while attractive female victims are more likely to be helped (Benson et al., 1976). These latter results, however, are found only amongst male bystanders (Benson et al., 1976). Such results may reveal the influence of physical appeal or feelings of attraction on

intervention behavior. The variable is scantly found in research, possibly due to the difficulty in controlling or measuring such a subjective characteristic. During an onboard assault, the deservedness/worthiness/social desirability effect may prove problematic for a flight attendant. Their prospect for intervention may depend on their attractiveness and whether a passenger believes the flight attendant (or even the airline) had provoked the attack through recent actions, attitudes, or policies (see Hunter, 2006). However, a high level of deservedness or worthiness does not always guarantee intervention.

In 1993, in Liverpool, England, two-year-old James Bulger was kidnapped, beaten, and murdered in public by two 10-year-old boys. Authorities confirmed that almost 60 adults observed the abuse, noticed distress and injuries on the toddler's face, and heard the child crying out for his mother (Rennie, 2021). However, none of the almost 60 adults intervened on behalf of the inherently innocent two-year-old victim (Christy & Voigt, 1994; Rennie, 2021). Formal research into aiding abused children has similarly found no more likelihood of intervention on behalf of an abused child than on behalf of an adult (Laner et al., 2001). Even a person recently being observed helping others does not have an increased chance that they will receive help (Allen, 1968). That revelation, in addition to the Bulger case, appears to show that making the victim more innocent or deserving does not guarantee a strong enough willingness on the part of bystanders to cause intervention. The more powerful influencer may be more complex and related to social interactions.

Social Interaction

Researchers have theorized that communication among bystanders would lead to more helping behavior (Latané & Nida, 1981). Results, however, showed the opposite—

when allowed to communicate with fellow bystanders, helping behavior *decreased* (Latané & Nida, 1981). The effect may not be directly related to communication, but conversing may instead be the method bystanders use to qualify the reactions of others. It may not be the mere presence of bystanders during an assault that inhibits helping behavior but rather their reactions and responses to it (Clark & Word, 1972; Latané & Darley, 1970).

Bystanders of an assault are part of a social event. Even if no verbal interaction occurs, individuals receive behavioral cues from others. In the context of intervention, bystanders who observe passivity in others are much more likely to remain passive as well (Latané & Darley, 1968; Latané & Rodin, 1969). The antecedent is also true. Witnessing another person intervening or exhibiting prosocial behavior increases the likelihood that a bystander will respond in turn (Bryan & London, 1970; Bryan & Test, 1967; Christy & Voigt, 1994; Latané & Rodin, 1969; Rushton & Campbell, 1977; see Clark & Word, 1972). The observing effect, however, has boundaries.

When bystanders are aware of others but cannot perceive their reaction, the collective pressure to remain passive increases (Darley & Latané, 1968). The pressure is presumably due to each bystander assuming that someone else is already responding (Darley & Latané, 1968). Written reactions may also be inadequate. During an internet chat room declaration of suicide, the classic bystander effect emerged and was not influenced by the written actions (either prosocial or passive) of others in the chat room (Orlando, 2020). However, if bystanders perceive others as incapable of helping, the lack of feedback becomes irrelevant; the bystander effect dissolves, and bystanders behave as

they would if they were alone (Bickman, 1972; Korte, 1971). The requirement for observable reactions extends to the victim as well.

A typical experimental design includes some indication of victim distress (often audible) in another room. Such designs have indeed resulted in participants displaying the classic bystander effect (Darley & Latané, 1968; Harris & Robinson, 1973; Latané & Darley, 1968; Ross, 1971; Schwartz & Gottlieb, 1976). However, if the victim is in *visible* distress, the bystander effect tends to be reduced (Ashton & Severy, 1976; Berkowitz & Daniels, 1963). The difference may be due to the ambiguity inherent in audible vs. visual signals. The lack of evidence-clarity becomes especially problematic for victims of online abuse, as the text-only medium reduces the visible and audible cues of their distress to zero (DeSmet, 2016). In an inflight environment, passengers can observe other passengers' actions and reactions (including passivity). The reaction and response of others, however, may not be what is influencing intervention. Instead, it may only be an antecedent to clarifying the event's severity.

Ambiguity

The interpretation of an event may dictate whether a bystander intervenes (Harada, 1985; Latane & Rodin, 1969; Shotland & Straw, 1976). If a bystander fails to interpret an event as one requiring intervention, then intervention will not occur (Latané & Darley, 1970). In events where the seriousness is undeniable, the bystander effect is absent (Clark & Word, 1972, 1974; Fischer et al., 2006; Philpot et al., 2019). Perhaps surprisingly, a reduction in the bystander effect is repeated even during extraordinarily violent and dangerous events, so long as the bystander can correctly interpret them as such (Fischer et al., 2006; Harari et al., 1985; Philpot et al., 2019). The seemingly

conflicting evidence indicates that the presence of others may be contributing less to the diffusion of responsibility and more to the vagueness of the event. An alternate explanation is that diffusion of responsibility is positively influenced by the level of ambiguity, though that theory is untested. As long as the ambiguity of help-needed is low, a reduction in the bystander effect also occurs in innocuous events, such as a person needing help with a bus or subway schedule (Allen, 1968; Harada, 1985) or an automobile breakdown (Hurley & Allen, 1974). Researchers have observed similar results in structured settings.

Controlled experiments reveal that the bystander effect diminishes in dangerous situations, even when intervention may result in negative consequences for the bystander (Fischer et al., 2006; Schwartz & Gottlieb, 1976). A possible explanation for the counterintuitive finding is that a violent assault is more readily and clearly interpreted as an emergency (low ambiguity) (Piliavin et al., 1969). However, such an explanation conflicts with the intervention evidence in the Genovese and Bulger murders. The evidence, however, supports findings that when many others act passively, the bystander effect becomes most pronounced.

Bystanders in large groups may assume that since nobody else perceives the need for intervention, the event must be benign (Latané & Nida, 1981). That interpretation account adds an alternative to the *diffusion of responsibility* explanation since a bystander interpreting an event as harmless will feel no level of personal responsibility to intervene. Alternatively, *both* reactions—correct interpretation and feeling responsible—are necessary prerequisites to a bystander deciding to intervene.

A flight attendant assault may be highly ambiguous. Passengers may misinterpret the event's seriousness due to other passengers failing to respond and other flight attendants failing to respond. Passengers may interpret the lack of a helping response on the part of the other flight attendants as evidence that the victim is handling the situation adequately. In such a case, a passenger-bystander would only respond if it provided some benefit to them.

The Self-Centered Bystander

Psychologists have attributed the actions of prosocial bystanders to an increased and distressing psychological arousal. The bystander acts (intervenes or flees) to reduce the arousal to below uncomfortable levels (Brewster & Tucker, 2016; Brigham, 1986; Piliavin et al., 1975; see Hortensius & deGelder, 2018). Researchers have repeatedly observed non-interveners showing physical signs of heightened arousal (trembling, sweating) (Darley & Latané, 1968) or reporting that they felt an increased level of distress (Batson et al., 1987; Eisenberg & Eggum, 2009; see Hortensius & deGelder, 2018). The evidence means intervening is less centered on *others* and more centered on the *self*. Of course, bystanders have the option of fleeing, which would also reduce arousal and is even more of a self-focused act. In adolescent bullying, child bystanders (particularly boys) often cope by physically distancing themselves from the victim (Hunter & Borg, 2006; Hunter et al., 2004; Pozzoli & Gini, 2010). It may be that for such young bystanders, fleeing is the only sure option for reducing feelings of distress.

These observations may indicate that even though people view intervention as a magnanimous act, it is, in reality, a self-centered one. The bystander is acting primarily to reduce their own level of distress instead of the victim's. During a flight attendant assault,

the option to flee is nonexistent. Being forced to remain may increase the physiological desire to intervene since it would be the only remaining method to reduce the feelings of distress.

The arousal may be dual-acting. When a person witnesses a victim engaged in a violent event, two competing motivations begin to build (Hortensius & deGelder, 2018). The first is a desire to either flee the scene (flight) or remain passive (freeze) (Hortensius & deGelder, 2018). Such a desire can manifest itself as anxiety or feelings of personal distress (Batson et al., 1987; Darley & Latané, 1968; Eisenberg & Eggum, 2009). The second motivation is a slower-building empathy-driven desire to help (Graziano & Habashi, 2010, 2015). When a person witnesses an emergency, the first motivation (flight or freeze) rapidly builds, which inhibits helping behavior (Hortensius & deGelder, 2018). As the slower empathy-driven motivation intensifies, it can overtake the first. If it does, helping will occur (Hortensius & deGelder, 2018). During an inflight assault, there is no opportunity for passenger-bystanders to act on the *flee* motivation. Even if the *freeze* motivation initially takes priority, there would be ample time (assuming an adequate duration of the assault) for the empathy-driven motivation to build and overtake the first. Unlike Earth-bound bystanders, passengers are forced to remain in the area while empathy-driven motivation builds. The sequence may result in a greater intervention rate during an inflight assault than elsewhere.

Location

Given that many attributed the Genovese murder to urban moral decay, researchers have attempted to verify the assumption that the bystander effect is a uniquely urban phenomenon. There is a correlation between the location of an incident

and bystander intervention (Levine et al., 1994; Milgram, 1970), with some verifying that the inhibiting effect is indeed stronger in cities than in rural areas (Gross, 1994; Latané & Nida, 1981; Merrens, 1973). However, even fundamental characteristics such as building size correlate with intervention (Newman, 1973). In all cases, however, researchers have stopped short of assigning any causality or explanation for the difference. Metadata confirms the correlation between intervention in private vs. public spaces (Hart & Miethe, 2008). However, those results are convoluted by many other variables (type of incident, weapon involvement, time of day, and victim-attacker relationship) (Hart & Miethe, 2008). Clues may come from college physical assault research, which revealed that students are more likely to intervene during an on-campus assault than an off-campus one (Brewster & Tucker, 2016). The difference may suggest that familiarity with a person's surroundings either enhances helping or inhibits remaining passive. Howard and Crano (1974) found similar results in those witnessing a theft. There were significant differences in reporting behavior in three locations (lounge, restaurant, and library) (Howard & Crano, 1974). The pair surmised that the level of formality of the location drove differences in intervention, though they had difficulty quantifying the "formality" variable. Their explanation may explain the on vs. off-campus difference, but researchers have avoided defining levels of formality or questioning participants on their perception of such.

Notwithstanding the gap in research surrounding location's effect, an assault on an aircraft occurs in a relatively formal and unfamiliar place. Passengers are assigned a specific seat, everyone faces the same direction, federal laws regulate behavior, there are clear lines of authority, and an aircraft is far different from any living or casual space.

Other location characteristics such as loud noises, densely crowded areas, obstructed views of an incident, and a bystander's preoccupation with other tasks correlate with reduced bystander intervention in varied scenarios (Fischer et al., 2011; Latané & Nida, 1981; see Burn, 2009; see Dovidio et al., 2017). All such characteristics are found on an aircraft and thus would be present during a flight attendant assault. Although researchers have yet to explain location's impact, what is clear is that bystander intervention varies based on where an incident occurs (Brewster & Tucker, 2016; Gross, 1994; Howard & Crano, 1974; Latané & Nida, 1981; Levine et al., 1994; Merrens, 1973; Milgram, 1970). The combination of unique qualities of an aircraft environment presents a gap in bystander intervention research. The inability to escape, social density, behavioral requirements, noise level, and limited view of an incident presents a combination of conditions not found elsewhere, much less in any existing bystander research.

Group Size

The epicenter of the bystander effect is the bystander group. Not surprisingly, researchers have examined this group to distinguish characteristics that would either enhance or reduce an individual's helping inhibition. Despite group size being the core discriminator underlying the bystander effect (see Latané & Darley, 1968), its full impact is far from conclusive.

A significant amount of early research seemed to solidify the impact of group size on helping (Latané & Darley, 1968, 1969, 1970; Latané & Nida, 1981; Latané & Rodin, 1969; Ross, 1971). Even during an innocuous event, such as knocking on a door, a large group seems to inhibit bystanders more than a small one (Levy et al., 1972). However, in a similar study, the time it took for a person to respond to knocking on a door *negatively*

correlated with group size (as group size increased, response time decreased), but only when there was a cost (continued distraction) to members of the group (Morgan, 1978). Removing the cost of nonintervention caused the group-size effect on response latency to disappear (Morgan, 1978). However, this is not always the case. A study of chat room requests for computer technical support showed that response latency was positive, albeit weakly, and correlated with the number of chat room members, even though the cost of nonintervention to bystanders was absent (Markey, 2000). These studies were unique by focusing on response latency and not simply whether someone helped or not.

Taken together, the lack of consistency in findings indicates that the effect of the group is not universal across all situations. The original conclusion that the group inhibited Genovese's neighbors from helping may be more of an event-unique finding than a general one. In fact, research into other violent street altercations found no main effect of group size on helping behavior (Levine & Crowther, 2008). Admittedly, the latter scenario differs significantly from knocking on a door or asking for computer support. Still, its finding is more meaningful because it closely relates to the Genovese case, the Bulger case, and an inflight flight attendant assault. However, the contrasting results in Levine and Crowther's (2008) research may be completely understandable and valid. After all, the victim in their study was not Genovese or Bulger, it was not 3:15 am, and it did not occur in New York or a crowded shopping center. The victim was also not a flight attendant, and the location was not an aircraft with dozens of bystanders—a combination absent in existing research.

Research has revealed a similar lack of group-size effect in cases of overly large groups, although the situations are very specific. Intervention research into stopping

someone from driving intoxicated showed that large group sizes (over 50) did not reduce levels of intervention (Rabow et al., 1990). Cyberbullying research involving over 5,000 bystanders found that the bystander effect only existed up to a group size of 24. There was no effect in larger groups (up to 5,000) (Obermaier et al., 2014). The reduction may mean that the diffusion of responsibility explanation (see Darley & Latané, 1968; see Latané & Nida, 1981; see Morgan, 1978; see Wallach et al., 1964) peaks at some amount but then declines as the number of bystanders continues to increase. During an assault onboard an aircraft, the number of bystanders can plausibly range over 100, easily half probably being close witnesses of the event. The number may exceed that required to achieve maximum bystander effect. Also, there appear to be higher-order effects of group size.

Chekroun and Brauer (2002) observed the classic bystander effect when someone vandalized an elevator but not when someone littered in a park. The nature of the offense and location may have combined with group size to alter bystander behavior, although Chekroun and Brauer (2002) admit several variables were left uncontrolled. Wiesenthal et al. (1983) observed the bystander effect during charitable giving, but different conditions resulted in differing group-size effects. In one circumstance, the group-size influence waned once the group reached only three members (Wiesenthal et al., 1983). Group size even affected giving when participants merely imagined the group.

Participants who imagined a group of 30 pledged less to charity than those who imagined a group of 10, who pledged less than those who imagined a group of only two (Garcia et al., 2002). The donation amount difference between the 10-group and two-group was also twice as much as the amount difference between the 10-group and 30-

group (Garcia et al., 2002). Such a nonlinear relationship is further evidence that the group-size effect peaks at some number, though Garcia et al. (2002) did not attempt to identify it. The research team replicated the results with participants imagining a movie theater either full or empty of all but a single friend. Garcia et al. (2002) then asked participants how much time they would be willing to commit to follow-on research. Interestingly, participants who imagined the group conditions committed significantly less time than those who imagined being with only a friend (Garcia et al., 2002).

These results show that research participants simply imagining being on a crowded airplane may exhibit the classic bystander effect when presented with an assault scenario. The results also show that the presence of others may have a lasting impact even when they are no longer present. The group may not directly affect behavior but may affect the bystander's persona, manifesting as reduced helping.

While being in a group of strangers increases the bystander effect, being in a group of friends does the opposite—it reduces the bystander effect (Levine & Crowther, 2008). The finding would add evidence that the group affects the bystander at a level deeper than behavior since peer pressure is far greater than that from strangers. Peer influence might be evident even if they have not been present for some time. However, distinguishing inconsequential peers from influential ones may prove difficult since the point at which one becomes the other is indistinct. When fellow college students comprised the group, the impact of group size (positive or negative) vanished (Levine & Crowther, 2008). These results may illustrate the latter point since fellow college students may not be *peer enough* to make a difference. Together, these results suggest that group

size alone is not the sole driving factor but is tempered by relational ties between the members.

Relationships Within the Triad

Relationships within the assault triad (victim, assailant, bystander) have caused mixed results. Relationship variants include bystander-bystander, bystander-assailant, assailant-victim, and victim-bystander. Experimental research in this area is limited, given the difficulty of recruiting large numbers of participants known to an experimental victim or bystander while maintaining research integrity. Researchers can use post facto analysis of real-world events to draw conclusions, but this approach is not without limitations since it only includes events reported to authorities.

Bystander-Bystander. The interdependency of actions between bystanders is at the core of the bystander effect. The theory, however, leaves out any impact of inter-bystander relationships. Research shows that bystanders being friends instead of strangers significantly reduces the bystander effect and response latency (Darley & Latané, 1968; Fischer et al., 2011; Latané & Rodin, 1969; Rabow et al., 1990). These results, however, are not conclusive. Flemish research into adolescent cyberbullying showed that friend-bystanders supported the victim *less* than acquaintance-bystanders, but only when the severity of the bullying was low (Bastiaenses et al., 2014). When the severity increased to a higher level, the results reversed—friend-bystanders helped *more* than acquaintance-bystanders (Bastiaenses et al., 2014). However, in this high-severity bullying scenario, friend-bystanders also supported the *bully* more than acquaintance-bystanders (Bastiaenses et al., 2014). In other words, friend-bystanders were likelier to help the victim *and* join in the bullying (Bastiaenses et al., 2014). Researchers observed similar

results during in-person adolescent bullying events, in which children are not only especially reluctant to intervene but often join in with the bullying of peers (Cunningham et al., 1998; O'Connell et al., 1999; Whitney & Smith, 1993). The wide variance shows that a bystander-bystander relationship is not distinct enough to be binary (such as friend/stranger). Instead, the relationship is a continuous-scale measure with different values for each person. Other scenarios have also yielded different results.

An analysis of intoxicated driving intervention showed that the number of other bystanders the intervener knows does not affect the intervening (Rabow et al., 1990). Even though Rabow et al. (1990) did not differentiate between friends and acquaintances, their results may indicate higher-order effects or more complex constructs are involved.

In more ambiguous situations, friend-bystanders may be apprehensive about being negatively judged if they act too impulsively by intervening when help is not needed (Latané & Darley, 1970). That apprehension may increase the bystander effect (Latané & Darley, 1970). High ambiguity may explain why friend-bystanders in the Flemish research were hesitant to exhibit prosocial behavior. Still, it does not explain the intoxicated driving results or why the Flemish friend-bystanders were more likely to support the cyberbully. It may be that the bystander-bystander relationship construct is more profound than simple friendship.

In a study involving response to a fire or injured person, bystanders responded more quickly when in the presence of children vs. adults (Ross, 1971). The difference may be evidence that a bystander-bystander relationship deeper than simple friendship is required to alter behavior. The relationship may be centered more on adherence to societal expectations. In the previous study, adult participants may have felt a social

obligation to protect children from fire. Even though other children-bystanders were present, the diffusion of responsibility equation did not include them, and thus, there was no sharing of responsibility. However, the above explanation does not align with the Bulger case, in which almost 60 adults failed to come to the aid of the toddler-victim. The explanation does align, however, if the diffusion of responsibility amongst the adults had a greater impact than the social expectation of the adults to protect the child-victim.

Bystander-Assailant. At the most basic level, research shows that a bystander who knows the assailant will tend to exhibit more victim-helping behavior (Christy & Voigt, 1994). The response assumes that bystanders disapprove of the assailant's actions, which may not always be accurate. Some segments of society approve of violence as a reaction to disrespectful or antagonistic behavior from the victim (Anderson, 2000; Miethe et al., 2004). That approval may affect helping, as behavioral expectations may substantially influence a bystander's behavior more than internal tendencies (Espelage et al., 2003; Gini, 2006, 2007; Juvonen & Galvan, 2008; Luckenbill, 1977; Miethe & Deibert, 2006). If those norms include passivity, it may heighten the bystander's inhibition to help. During any abnormal event on an aircraft, the expected behavior is to "Remain seated with your seatbelts fastened." Passengers hear it before takeoff, when in turbulence, and during an emergency. The flight crew wants you to stay in your seat if an abnormal event occurs (like a physical altercation). A passenger getting up to help a flight attendant requires first breaking this behavioral norm. Essentially, a passenger-bystander must willfully disobey the flight attendant before assisting them. The requirement may cause a strengthening of the bystander effect.

Bystanders may also take cues from the treatment an assailant has recently received. Students who witness tacit approval from a teacher toward an aggressive student may assign value to the aggressive behavior (Chang, 2003). Doing so may make students more likely to judge future aggressive behavior as acceptable (Chang, 2003; see Coie & Koeppl, 1990). Therein lies a potential dilemma for some passengers since they may see the flight attendant acting courteously, politely, and respectfully to the assailant beforehand. These episodes may create a preamble for how a passenger-bystander interprets the subsequent assailant-flight attendant conflict.

Assailant-Victim. If a bystander assumes an assailant and victim are related, they are less likely to intervene (Laner et al., 2001; Levine et al., 2002; Shotland & Straw, 1976). Using data from actual encounters, however, shows the relationship to be more complex and highly dependent on second and third-order effects of location, type of crime, presence of a weapon, and time of day (Hart & Miethe, 2008).

In general, bystanders are less likely to intervene if they know a relationship (husband-wife, parent-child, dog-owner) exists between the attacker and victim (Laner et al., 2001). However, the tendency to remain passive may not be directly due to the perceived relationship, but instead due to the ambiguity the relationship creates. A person witnessing a conflict between two related individuals may assume they only see a small portion of an enduring and more deep-rooted conflict. The assumption may lead to uncertainty concerning the victim's deservedness, the aggressor's intent, or even which person is in which role. A lack of relationship clarity may explain the passivity of the Bulger adults since the victim and assailants were all preadolescent children—a dynamic not found in adult bystander literature.

Victim-Bystander. Researchers generally agree that individuals are more inclined to assist a friend than a stranger (DeSmet, 2016; Leone et al., 2020; Levine et al., 2002), but the discriminator between stranger and friend is undefined. In school bullying events, a victim's peers are present in as much as 85% of incidents (Atlas & Pepler, 1998; Craig & Pepler, 1995), though intervention rarely occurs (Hawkins et al., 2001). It is also not uncommon for peers to join in the bullying (Cunningham et al., 1998; O'Connell et al., 1999; Whitney & Smith, 1993). One-on-one contact between bystander and victim, either during the event or shortly beforehand, seems to consistently reduce or even eliminate the bystander effect (Howard et al., 1974; Markey, 2000). It may be that once personal contact occurs, responsibility is no longer diffused among bystanders. The role and responsibility of the targeted bystander are now unique from all others present.

Introduction of Constructs

The differing effects of observed variables have led to suspicions that the true causes of intervention and passivity are much more profound. Researchers have examined the impacts of latent constructs for clues to intervention behavior. Understandably, they were slow to do so since the earliest pioneers of the bystander phenomenon found no link between personality traits and bystander behavior (see Darley & Latané, 1968). More recent research, however, has identified such correlations (Graziano & Habashi, 2015). Constructs such as empathy, awareness, attitudes (Nickerson et al., 2014), aggressiveness (Laner et al., 2001), sympathy (Batson et al., 1987; Eisenberg & Eggum, 2009; Huston & Korte, 1976; Laner et al., 2001), personality (Eisenberg et al., 2002; Karakashian et al., 2006; Latané & Nida, 1981; Michelini et al., 1975; Zoccola et al., 2011), shyness (Karakashian et al., 2006; Latané & Nida, 1981;

Zoccola et al., 2011), and moral judgment (Eisenberg et al., 2002; Huston & Korte, 1976) have all been strong predictors of intervention. However, these predictors are inconsistent and often vary in the second order. For example, neither benevolent sexism (sexism beneficial to the opposite sex, such as a male's duty to protect and care for women) nor social status correlated with bystander behavior, but the combined effect was significant (Leone et al., 2020). However, these results were confirmed when the victim was a friend but not when the victim was a stranger (Leone et al., 2020). Such specificity is common amongst many of the above constructs. Since most of the results are specific to only a single scenario, the question remains about how much influence they have on a bystander's intervention.

Types of Intervention

Though none of the witnesses tried physically to stop Genovese's attacker, one person did yell out, "Leave that girl alone!" (Krajicek, 2011). Others may also have called the police, although reports are unclear (Ruhl, 2021). Both types of indirect assistance still qualify as helping behavior.

Witnesses to an attack have a choice of four behaviors: joining the attacker (Dillon et al., 2015), remaining passive, direct intervention (such as pulling the attacker away), or indirect intervention (such as calling the police or telling a teacher) (Dillon & Bushman, 2015; Hart, & Miethe, 2008). The former two are enabling behaviors, while the latter two are prosocial behaviors (see Batson, 1998; see Salmivalli, 2010).

Although many consider passive bystanding a neutral behavior (Salmivalli & Voeten, 2004), it is, in fact, an enabling act. It provides feedback that interference will not be forthcoming, thus silently approving the attack (Cowie, 2000; Kowalski et al.,

2014; Menesini et al., 2003; Salmivalli, 2010; Salmivalli et al., 1996, 2011). Introducing video capability (smartphones) presents a new and indistinct behavior not found in bystander research.

Videoing an assault may encourage the attacker since it conveys evidence that what is occurring is entertaining. However, a bystander may video for unspoken but prosocial reasons. They may feel unable to intervene but decide they can help by capturing video proof to be used for later punitive action. However, whether this possibility acts as a deterrent to continuing an attack is unresearched. There is also a research gap concerning video effects on other bystanders performing indirect or direct intervention.

Research into self-reported intervention by college students showed that the variables *on-campus*, *off-campus*, and *state of hurriedness* did not affect whether intervention occurred but did affect whether the participant intervened directly or indirectly (Brewster & Tucker, 2016). In cyberbullying, indirect intervention occurs significantly more frequently than direct intervention, which in this case involves directly addressing either the bully or victim (Dillon & Bushman, 2015). During an assault on a flight attendant, the direct intervention methods consist of separating the attacker from the victim or restraining the attacker. Indirect intervention options are limited but can include calling for other flight attendants or passengers to intervene.

Recent Contributions

Traditional bystander research has focused on the singular bystander and their actions. Researchers often design their study to measure the impact of one or more variables on a single person's actions. The most frequently included variable is group size

(Banyard, 2008). Often, other group members are confederates who have been instructed *not* to respond. That strategy isolates the participant's performance. Classic bystander results lead people to assume that if attacked, it is better if only one person is present than many (Levine et al., 2020). Such assumptions, however, may be false because they are based on the measured response of a single bystander.

One of the pioneers of bystander research recognized this problem, writing, "It is meaningless to compare directly individual with group responses, since with differing numbers of people available to respond, there is a purely mechanical potential for getting more help with more people" (Latané, 1981, p. 350). Taking too much of a bystander-centric approach to violent events ignores the real problem—that a *victim* is being attacked. In every case, the victim's primary concern is whether someone—anyone—will come to their aid (Latané & Nida, 1981). The misdirection is partially due to the statistical conclusions drawn from situations with especially large numbers of bystanders.

As the number of helpers increases, the perceived amount of help required from the remaining bystanders decreases. The amount peaks at around three helpers (Levine et al., 2011). In other words, if three people are already helping, most other bystanders assume that adequate help is being rendered (Liebst et al., 2019). Researchers, however, will observe only three out of 100 bystanders assisting and conclude that 97% were unwilling to help. Based on the evidence from Liebst et al. (2019), such conclusions may be false. Research in the last decade has examined the bystander effect less from the bystander's perspective and more from the victim's.

A review of closed-circuit television (CCTV) footage of actual severely violent events illustrates the dichotomy of result analysis. From a purely bystander-centric

approach, the classic bystander effect is apparent, meaning the number of bystanders negatively correlates with the chance that each bystander will intervene (Liebst, 2019). However, a victim-centric view of the same footage shows that additional bystanders increase the chances that one or even several people will intervene—even during dangerous encounters (Fischer et al., 2011; Levine et al., 2020; Liebst, 2019; Philpot, 2019; Philpot et al., 2020). Similar CCTV analysis shows that each additional bystander *increases* the odds that the victim will receive help (Levine et al., 2011; Philpot et al., 2019, 2020), and that bystander intervention during violent encounters in public is the more normative behavior (Philpot et al., 2019). The evidence is understandable when considering that the bystander effect is likely nonlinear. In other words, the difference in intervention rates between two and three bystanders is greater than between 12 and 13 (see Liebst, 2019). Those results may mean that in a scenario with many bystanders (like an aircraft), the group's effect may *level off*. As a result, the chance of a flight attendant receiving help from someone—anyone—may be greater than previously assumed.

The approach of using CCTV footage as data is unique because it is unbounded from using only reported incidents (Philpot et al., 2020). Researchers who utilize this approach use data from real life as it occurs (Philpot et al., 2020). However, they also introduce the camera as an uncontrolled but potentially significant variable. CCTV evidence alone does not indicate the behavioral effect of noticing the camera, or even if any in the triad did so. Also, CCTV footage restricts the viewer to only seeing what happens within the camera's field of view. It does not capture any preceding social interaction that occurred elsewhere (Liebst et al., 2019; Philpot et al., 2020).

Recreating these incredibly violent encounters is difficult in the experimental setting. However, the few that have been successful confirm that a bystander alone is just as likely or even more likely to intervene than when others are nearby (Fischer et al., 2006; Harari et al., 1985), even though the threat to themselves is higher when alone. This contrary evidence means researching the bystander effect across widely varied scenarios is paramount.

Social-Structure. Only in the last decade have researchers more closely examined the complex social structures within a violent assault. Results indicate that the effects of social relationships between the victim, assailant, and bystander were greater than the group size (Levine & Manning, 2013; Liebst et al., 2019). The key to understanding bystander behavior possibly lies more in the relationships within the triad than merely in group size (Levine & Manning, 2013; Liebst et al., 2019). The understanding means that when considering a victim's plight, the presence of others is not the obstacle that needs to be overcome (Levine et al., 2020). The real obstacle is misunderstanding the relational complexities within the triad of event actors. Practitioners may serve victims better by considering the group not as a threat but as a source of intervention power waiting to be exploited. (Levine et al., 2020).

The Bystander Intervention Model

Bystander research primarily focuses on the various situational details and personal qualities that amplify or suppress intervention behavior. Researchers usually restrict response measures to a binary option (help, not help), particularly in experimental designs (see Brewster & Tucker, 2016; see Darley & Latané, 1968; see Dillon & Bushman, 2015; see Fischer et al., 2006; see Greitemeyer et al., 2006; see Grzyb, 2016;

see Harris & Robinson, 1973; see Hurley & Allen, 1974; see Latane´ & Nida, 1981). Such a strategy is convenient for those only concerned with whether intervention occurs, but it tends to ignore the triggers that lead to helping and the barriers that block it. Filling the gap requires exploring the cognitive process that precedes the response.

Bystanders who intervene in an emergency, such as an assault, must notice the event, recognize the need for help, and accept personal responsibility to assist (Latané & Darley, 1968). Bystanders must also know what to do before deciding to take action (Latané & Darley, 1970). Pioneers of bystander research modeled the five steps to explain the process that leads to intervention (see Figure 2) (Latané & Darley, 1968, 1970; Rabow et al., 1990; see Dovidio et al., 2017). Failure to accomplish any of the steps results in passivity.

Figure 2

The Bystander Intervention Model

Note. BIM steps taken from *The Unresponsive Bystander: Why Doesn't He Help?* by B. Latané and J. M. Darley, 1970, Appleton-Century-Crofts.

Bystander theory holds that each step of the bystander intervention model influences the accomplishment of its following action (Albayrak-Aydemir & Gleibs, 2021; Nickerson et al., 2014; Rabow et al., 1990; see Anker & Feeley, 2011; see Christy & Voigt, 1994). Researchers attempted to combine the five steps into a single construct

but found results inconsistent with the data (poor model fit) (Aydemir & Gleibs, 2021; Nickerson et al., 2014). The inconsistency revealed the failure of the single construct to capture all the dimensions of a bystander's decision to help. It also illustrates that a single inhibitor may not be what drives passivity. Instead, many inhibitors may act on one or more of the dimensions. Contrary to popular understanding, bystander apathy is not truly what inhibits helping. Instead, the influence of personal and situational barriers blocks the bystander from fulfilling one of the model's five steps (Anker & Feeley, 2011; Latané & Darley, 1968, 1970; Latané & Nida, 1981).

Noticing the Event

Before a bystander can intervene, they must notice that the event is happening. Some events readily command attention, such as when Kitty Genovese screamed, "Oh, my God, he stabbed me! Please help me!" (Krajicek, 2011). Other events, however, are nearly void of cues. During the first minutes of the Bulger case, people merely saw two 10-year-old boys leading a younger boy—one holding his hand—through a crowded shopping mall (Rennie, 2021). To those present, nothing suggested that what they witnessed was a kidnapping.

James Bulger did not readily appear to be a victim. The absence of such signals makes it difficult for onlookers to discern a person at risk. (Small & Loewenstein, 2003). During his kidnapping, people were also preoccupied with shopping and immersed in an environment filled with noise, crowds, and distractions. These elements make it even less likely that an event will capture a bystander's attention (Fischer et al., 2011; Latané & Nida, 1981; see Burn, 2009; see Dovidio et al., 2017). During a flight attendant assault, noise, obstructed view, other passengers, and preoccupation with electronic displays may

inhibit a passenger from noticing the incident. Those who do notice must recognize the need for help before intervening.

Recognizing the Need for Intervention

If a person does not need help, there is no benefit to offering it. A bystander may notice something is happening but misinterpret it as *not* requiring intervention (Greitemeyer et al., 2006). The error is often present in sexual assaults (Burn, 2009; Labhardt et al., 2017) and racial violence (Nelson et al., 2011), where passive bystanders commonly report uncertainty about what they have seen and ultimately mischaracterize the event as harmless. In the Bulger case, some adults did inquire about the injuries to the toddler's face but, upon hearing a plausible explanation, were satisfied that no further action was needed (Rennie, 2021). Additional bystanders create added barriers to recognition.

Bystanders may rely heavily on the overt actions of others in determining an event's seriousness (Darley & Latané, 1968; Latané & Darley, 1970). Collective passive behavior can result from pluralistic ignorance, in which nobody helps because it looks like no one else thinks help is needed (Burn, 2009). Each bystander may be uncertain about what they are seeing and uses the aggregate hesitancy of others as confirmation that the event must be harmless. Pluralistic ignorance may have contributed to so many adults failing to recognize the need for help during the latter stages of the Bulger case. The collective passivity may have convinced each person that the child was safe despite evidence to the contrary.

Taking Responsibility

After a bystander recognizes the need for help, they must overcome any diffusion of responsibility (Burn, 2009) and internalize a level of personal obligation more significant than that ascribed to the group (see Wallach et al., 1964). In bullying cases, this sense of personal responsibility positively correlates with intervening and negatively correlates with remaining passive (Pozzoli & Gini, 2010). The same correlation exists in child abuse cases (Banyard, 2008; Christy & Voigt, 1994) as well as both actual (Banyard et al., 2007, 2014; Burn, 2009) and virtual-reality (Jouriles et al., 2016) violent conflicts. Even in simple circumstances such as recycling, accepting personal responsibility is a predictor of follow-on model steps (Anker et al., 2011). However, the responsibility can have many targets. The traditional view is victim-centric, focusing on the responsibility to the victim. That view is readily apparent in the research on bullying, child abuse, and violent conflicts but becomes less evident in the recycling study. The victim, in this case, may vary across bystanders. Some might see the Earth, society, or children as the victims of nonrecyclable waste. Others may even view themselves as the primary victim affected by waste, making their actions no longer those of a bystander. These and other perceptions of the victim and attacker can also become barriers to accomplishing this step.

Bystanders may absolve themselves of responsibility if they believe a victim's predicament is the result of their own choices or carelessness (Burn, 2009; Loewenstein & Small, 2007). Perceived personal relationships between victim and assailant can also hinder bystanders at this step (Burn, 2009; Laner et al., 2001; Levine et al., 2002; Shotland & Straw, 1976) since they may feel it is not their responsibility to mediate a

domestic dispute (see Laner et al., 2001). Accepting responsibility leads to a multi-option step of the intervention model.

Knowing What to do

Before bystanders can help, they must know what actions will effectively achieve positive outcomes. Step four may be the most conscious of the five, demonstrated by bystanders occasionally verbalizing their dilemma (i.e., "What should I do?") (Darley & Latané, 1968; DeCarlo, 1988). Failing to know what the situation requires (Hazler, 1996; see Banyard, 2008) or believing that one lacks the ability (Burn, 2009) are barriers to a bystander knowing what they individually should do. Step four is particularly problematic during adolescent bullying since many child-bystanders do not have the life experience necessary to know how to respond sufficiently (see Cunningham et al., 1998; see O'Connell et al., 1999; see Salmivalli et al., 1996; see Whitney & Smith, 1993). However, this step is unique because it is the only one where bystanders have options.

Bystanders can choose between many direct and indirect options when considering how to respond. Unlike previous steps in the model, a bystander at this step may be more deliberate instead of reflexive. They may consciously undergo a cycle of selecting, evaluating, rejecting, and reselecting various intervention options. The Genovese neighbor who called out, "Leave that girl alone!" may have done so only after consciously rejecting the physical intervention option. Selecting any option, direct or indirect, positions a bystander for the final model step.

Deciding to Act

The ending step of the bystander intervention model is deciding to act. Some research designs measure only this step, asking a variation of the question, "How willing

are you to help?" An affirmative response indicates that the participant accomplished the previous four steps, but a negative response gives no indication of where the inhibition occurred. A barrier somewhere impacted the participant, but its location could be at any of the steps, even this final one.

Bystanders may hesitate at the final stage due to apprehension about being ridiculed or poorly judged for their actions (Burn, 2009; Latane & Darley, 1970). Accomplishing step five requires either an alignment with behavioral norms (Burn, 2009; see Luckenbill, 1977; see Miethe & Deibert, 2006) or the ability of the bystander to ignore any norms that may run counter to their plan of action (such as the expectation to remain seated). In the absence of any of these barriers, an intervention will occur.

Bystander Intervention Model in Context

Few researchers have evaluated the applicability of the full bystander intervention model. Those who have done so report varied results across contexts. Full-model estimations (all five steps) in bullying (Nickerson et al., 2014) and sexual assault incidents (Nickerson et al., 2014; see Burn, 2009) achieved good model fit and validated the premise that each step predicts its following action. Anker and Feeley (2011) also conducted model estimations of the first four steps in the contexts of organ donors and green living (recycling, waste reduction). Although their model excluded the final step (deciding to act), it replicated the results found in prior research (see Burn, 2009; see Nickerson et al., 2014). Their results showed that those who perform prosocial behavior score higher on all four steps than those who do not (Anker & Feeley, 2011). However, results are not uniform across all scenarios.

Estimating the five-step model for charity donations shows *knowing how to help* (step four) is *not* a significant predictor of *deciding to act* (step five) (Albayrak-Aydemir & Gleibs, 2021). Results were identical for providing political support to refugees (Albayrak-Aydemir & Gleibs, 2021). The inability of either model to predict the last step may be due to the absence of immediate feedback from the participant's actions. The benefits gained from donating to charity or supporting a political candidate are not apparent until much later. Model estimation also differed across nationalities (Albayrak-Aydemir & Gleibs, 2021), further indicating that the model's applicability may have limits in terms of context and population.

These limits highlight the importance of utilizing the entire model in bystander research instead of only the final step (*deciding to act*). Variables and constructs may influence different steps of the model under different contexts. Perceptions of Genovese—an adult female on the street at night—may have influenced the bystanders' *taking responsibility* (step three). In contrast, the perceptions of Bulger—a small child with other children—may have influenced the bystanders' *noticing of the event* (step one). Both cases include unique victim perceptions, but by only applying the final step of the model, any difference in perception effects becomes lost. In other words, researchers may discover perception affects the decision to act, but they won't know how. Additional non-model-centric research designs have used the steps as variables.

Kalafat et al. (1993) examined the effects of ambiguity and group size on steps two and three (*recognizing the need for intervention* and *taking responsibility*) of the model, as well as the ability of these steps to predict prosocial child-abuse intervention behavior. Their results aligned with traditional findings of ambiguity and group size

effects on helping. Moreover, they clarified how bystanders internalize ambiguity and group size by deviating from the conventional binary response (help, not help) and using the two model steps as dependent variables instead. The two steps also correlated positively with child abuse intervention (Kalafat et al., 1993). The correlation supports, at least in part, the model's applicability to child abuse intervention.

Hoefnagels and Zwikker (2001) used a similar approach in their child abuse research by examining the first two steps of the intervention model (*noticing the event* and *recognizing the need for intervention*). Like Kalafat et al. (1993), the results were positive but only partially supported the model's applicability in child abuse cases since the research pair excluded model steps three through five. The value of Kalafat et al. (1993) and Hoefnagels and Zwikker's (2001) research is that it creates an understanding that as a bystander proceeds through the steps of the model, one or more external factors influence them along the way. The final decision to act may be less a direct result of these factors and more of a byproduct of their influence during the previous four steps.

Even though the model's limits and applicability are expanding, there are still significant gaps. There is no research using the entire model in the context of a physical assault or any incident on an aircraft. There is also no research examining exogenous predictors of the whole model.

Nonintervention Cost

The first three hijacked flights on 9-11 impacted their targets with passengers still passively seated, believing the flights would return to the airport (Janos, 2020). However, the passengers on United Flight 93, aware of the first three outcomes, knew they would surely perish if they did not act (Janos, 2020). Their risk was in *not* intervening.

In most research, the structure of a conflict is such that the bystander incurs the cost of intervening (see Allen, 1968; see Fischer & Greitemeyer, 2013; see Latané & Darley, 1970; see Latané & Nida, 1981), while the victim incurs the cost of *not* intervening (see Clark & Word, 1972; see Fischer et al., 2006; see Hart & Miethe, 2008). Research scenarios in which the bystander would suffer from nonintervention are rare (see Austin, 1979; see Latané & Darley, 1968; see Morgan, 1978), and researchers have yet to vary the nonintervention cost. The gap means the impact of this unique cost is unknown. At best, readers can conclude that any negativity incurred through passivity might suppress the influence of other variables such as race (Gaertner & Dovidio, 1977; Gaertner et al., 1982) or gender (Eagley & Crowley, 1986), but the gap in research makes even this speculative. However, if the assumption proves valid, it may indicate an internal evaluative process in which a bystander creates a complex 2x2x2 matrix of risks and rewards from both intervention and nonintervention to the victim and themselves. Bystanders would then select the optimum combination and act accordingly (see Austin, 1979; see Piliavin et al., 1981). The process may create a unique change in the perceived role. At some point, the cost of nonintervention may become so high that a witness begins to see themselves less as a bystander and more as a directly-affected co-victim. Such a shift would manifest itself in the bystander assigning a higher level of seriousness to the event, thus diminishing the influence to remain passive.

Many passengers on Flight 93 realized the fatal cost of nonintervention. For those passengers, influences like those driving passivity on the other three flights vanished. They perceived the event as severe enough that the crew needed help. During an assault on a flight attendant, risks of nonintervention are much less than that on Flight 93, but

they may still be present. If the confrontation continues, the safety and emotional wellbeing of passengers may be threatened (Bell, 2022), and there may be an additional risk of disrupted travel plans, costs due to delays, or even personal injury. These possibilities may make bystanders more apt to evaluate the assault as severe enough to require help.

Out-Group Social Bias

Every person belongs to one or more social groups (Montagu, 1949). The groups an individual belongs to are *in-groups*, while all others are *out-groups*. A group's shared traits become part of each member's identity (Allen, 1968; Nesdale & Scarlett, 2004; Ojala & Nesdale, 2004). When people evaluate a person, they utilize, to varying degrees, the shared traits (real or imagined) of the out-group to which the other person belongs. Observers may define an individual more by the attributes of the out-group than by details of the individual (Aron et al., 1992; Sedikides et al., 1993; Smith & Henry, 1996; see Cameron et al., 2006; see Tajfel & Turner, 1979). While the bias toward one's own group is almost universally positive, the bias toward out-group members is almost universally negative (see Abbott & Cameron, 2014; see Brewer, 2001; see Brewer, 1979; see Doise et al., 1972; see Gardham & Brown, 2001; see Hewstone et al., 2002; see Levine et al., 2005, 2020; see Rabbie & Horwitz, 1969; see Stephenson et al., 1976; see Tajfel, 1979, 1982; see Tajfel et al., 1971; see Wilson & Miller, 1961). However, the distinction perpetuates the narrow assumption that bias is an inherently negative impulse (see Hewstone et al., 2002). The bias between groups can indeed cause people to develop different treatment standards for others, but they are not always predisposed toward disfavor.

Treatment bias can take different forms. It has led to both increased helping (Berkowitz, 1972; Berkowitz & Daniels, 1963; Cialdini et al., 1997; Darley & Latané, 1968; Dovidio et al., 1991; Emswiller et al., 1971; Gini, 2006; Gottlieb & Carver 1980; Howard & Crano 1974; Levine et al., 2002; Levine & Crowther, 2008; Rabow et al., 1990; Rutkowski et al., 1983), increased passivity (Cowie, 2000; Hawkins et al., 2001), as well as increased hostility (Black, 1993; Levine et al., 2012; Phillips & Cooney, 2005; Swann et al., 2010) depending on the situation and perceptions of the groups. Even very superficial categorization of groups can have a direct effect on treatment bias toward its members (Brewer, 1979; Doise et al., 1972; Gardham & Brown, 2001; Levine et al., 2002, 2020; Rabbie & Horwitz, 1969; Stephenson et al., 1976; Tajfel, 1982; Tajfel et al., 1971). In an aircraft cabin, a person may not have any relationship or shared traits with fellow passengers, but the mere fact that they are *not* flight attendants creates a clear distinction between the two groups. However, such a distinction does not presuppose animosity.

Allegiances toward one's own group do not necessarily equate to an equal but opposite degree of discord toward outsiders. In terms of allocating assistance, a positive bias toward one's own group is more influential than a negative bias toward outsiders (Blanz et al., 1995, 1997; Gardham & Brown, 2001; Mummendey et al., 1992; Mummendey & Otten, 1988; Otten et al., 1996; see Vanbeselaere, 1987). The difference may indicate a more substantial influence of positivity, even toward outsiders, although that possibility is unresearched. Research instead has focused on factors that may alter levels of treatment bias.

A person's bias toward an out-group can be highly influenced by their demographics (age, gender, income) (Dion, 1973; Gini, 2007; Katz, 1976; Wilson & Kayatani, 1968; see Aboud, 1988; see Fazio et al., 1995; see Plant & Devine, 1998; see Rutland, 1999; see Rutland et al., 2005.) and the perceived status of the outside group (Bettencourt et al., 2001; Brewer & Brown, 1998; Brewer and Campbell, 1976; Duckitt, 1992; Ellemers et al., 1993; Gini, 2007; Hagendoorn, 1995; Harris & Robinson, 1973; Jost, 2001; Midlarsky, 1971; Midlarsky and Midlarsky, 1973; Mullen et al., 1992; Reichl, 1997; Taifel, 1982; see Boldry & Kashy, 1999). While the influence of demographics is unremarkable, the relevance of status is notable.

Research into the effect of group status on helping is minimal. The few studies which report the variance appear to imply that presumptions of bias between high and low-status groups being inherently negative may be unfounded. Although not directly addressed, results suggest high-status groups may experience positive treatment bias from low-status groups in some circumstances (see Gini, 2007; see Harris & Robinson, 1973; see Nesdale & Scarlett, 2004), particularly if the distinction is perceived to be justified (see Boldry & Kashy, 1999; see Duckitt, 1992; see Jost, 2001). However, direct effects of status, its different categories (economic, power, education), or scenarios involving positive bias toward out-groups are unresearched.

Due to the presumption that all bias is unfavorable, research has ignored any evaluation of positive bias directed toward an out-group member. If perceptions of negative attributes about a group transfer to a person's assessment of its members, then the same might be true of positive attributes. If observers perceive traits of an out-group to be of higher quality, it might create a bias toward favoring its members. Passengers

expect flight attendants to portray high-quality characteristics (friendly, helpful, competent, well-groomed, etc.). There are also clear distinctions between passengers and flight attendants regarding authority, responsibilities, and roles. Few passengers would argue that these differences are unjust, even though they clearly establish flight attendants as the more socially powerful group. Although traditional research has only focused on the impact of negativity toward out-group members, the positive attributes assigned to flight attendants may equate to a positive bias toward helping.

Intervention Skills

If bystanders do not believe they have the skills necessary to affect change, they will not attempt to do so (Burn, 2009; Cramer et al., 1988). The impact of various types of training on a bystander's evaluation of their abilities varies widely. Results show some training has a positive effect (Banyard, 2007; Banyard et al., 2007; Brewster & Tucker, 2016; Clark & Word, 1974; Cramer et al., 1988; Huston et al., 1981; Laner et al., 2001; Pantin & Carver, 1982; Potts & Lynch, 2010; Shotland & Heinold, 1985) and others have no effect (Brewster & Tucker, 2016; Huston et al., 1981; Laner et al., 2001). Evaluating the impact of acquired skills becomes complex since they are only one element influencing a person’s perception of their contribution’s worth.

The perceived abilities of other bystanders (Cramer et al., 1988; Darley & Latané, 1968; Horowitz, 1971; Piliavin et al., 1975) and past experiences (Banyard, 2008; Huston et al., 1981; Laner et al., 2001) can alter the level of value a bystander assigns to their own skills. A bystander measures their abilities against those required by the situation, but they may perceive the conflict to be uniquely different from the ones they experienced in training. When a person assesses a violent event, they may assign

different levels of relevance to the abilities of the triad members, the environment in which it occurs, and their own moral and ethical responsibilities (see Huston et al., 1981). The differing levels of relevance may mean a bystander is unsure about what to do despite receiving related intervention training. In the case of a flight attendant assault, a passenger may assume that the training allocated to flight attendants exceeds any that they have received. Their assumption may reduce the belief that their skills would be of any further assistance. Alternatively, a passenger may view the scenario as precisely matching what they have been trained for, which would result in assigning a high value to their skill, increasing their belief that they know how to respond. When an American Airlines flight attendant was attacked on a flight to New York, it was an off-duty police officer who was the only one to intervene—restraining the attacker until the flight landed (WABC, 2021).

Expectation of Positive Outcome

When a person intervenes, they expect their actions to have specific results (Bandura, 1977, 1986). They wish to achieve the desired end state, but it is often unclear what that state is. Empathy granted to the victim (Sainio et al., 2011) or confronting deviant behavior to stop a situation from occurring (Chekroun & Brauer, 2002; see O'Connell et al., 1999; see Salmivalli et al., 2011) may influence a person's decision to intervene. If a bystander does not expect their actions to support at least one of these outcomes, the intervention will not occur.

Victim-Centric Intervention

A victim-centric approach to intervention centers on comforting and making the victim feel better (Sainio et al., 2011). In essence, the helper's objective is focused less

on stopping the event and more on reducing its impact on the victim. Not surprisingly, pro-victim feelings correlate with greater levels of helping behavior (Davis, 1994; DeSmet et al., 2016; Eisenberg, 1991; Eisenberg & Fabes, 1998; Endresen & Olweus, 2001; Gini et al., 2007, 2008; Hoffman, 2001; Philpot et al., 2020; Pöyhönen et al., 2012; Rudolph et al., 2004; Weiner, 1980). Victim-centric helping is particularly prominent in violent offenses (Hart & Miethe, 2008), although bystanders may reduce their risk by delaying assistance until after the assault ends. While nobody physically helped Genovese as she was attacked, one of her neighbors emerged afterward and held her as she lay dying.

If passengers take a victim-centric view of a flight attendant assault, it may positively influence their decision to act. However, a victim-centric attitude can be problematic if bystanders expect their intervention will worsen the victim's predicament. Paradoxically, a bystander may feel that the best way to help a victim is not to become involved (see Goffman, 1959; see Lofland, 1969; see Miethe & Deibert, 2006; see Pöyhönen et al., 2012). A passenger-bystander may want to help a flight-attendant victim while fearing their intervention will agitate the attacker. The latter may cause the passenger to decide that remaining seated is the best strategy for preventing the assault from becoming more serious.

Situation-Centric Intervention

Sometimes, bystanders may focus their intervention efforts more on stopping objectionable behavior. Research into that focus is limited and has so far been restricted to online and in-person bullying. Bystanders who exhibit situation-centric intervention focus more on stopping the event—for example, by chastising and reprimanding the

offender—than on comforting and protecting the victim. As expected, bystanders who believe they can stop the aggressor's behavior are more likely to intervene (Dillon & Bushman, 2015; Pöyhönen et al., 2012). In cyberbullying, situation-centric helping is notably more prominent than victim-centric helping (Dillon & Bushman, 2015). The reasons may be two-fold. The inherent distance from both victim and attacker reduces the opportunity to provide comfort to the victim and the risk of physical confrontation with the aggressor. Such distance is, of course, absent on an aircraft, but the formal and well-defined behavioral expectations may elicit a strong passenger desire to rebuke anyone violating the norm.

General Self-Efficacy

General self-efficacy is an individual's belief in their ability to perform effectively (Bandura, 1997). It is not a measure of acquired skill but rather a self-perception of general ability based on the combination and interactions of a person's personality, life experiences, moral beliefs, and personal development. People with high general self-efficacy undertake more challenging tasks (Bandura, 1997), believing they can succeed even when faced with a new experience. Many bystander intervention researchers slightly narrow general self-efficacy to the boundaries of their targeted scenario, but most will still measure it in general terms, such as *violence* or *intervening*. However, the use of true general self-efficacy in intervention research is absent.

When faced with an interpersonal conflict, bystanders exhibiting high self-efficacy are much more likely to intervene (Caprara & Steca, 2005; Caprara et al., 2003; DeSmet et al., 2016; Gini et al., 2008; Pöyhönen et al., 2010; Pöyhönen & Salmivalli, 2008; Pöyhönen et al., 2012; Schwarzer et al., 1992; Thornberg et al., 2012), even in

violent encounters (Banyard, 2008). Before conceptualizing self-efficacy in 1986, researchers used the terms *competence* and *confidence* but still found them positively correlated with helping behavior (Ashton & Severy, 1976; Kazdin & Bryan, 1971; Midlarsky, 1971; Midlarsky & Midlarsky, 1973; Schwartz & David, 1976). Even perceived abilities that are only very peripherally related to the type of event positively correlate with helping (Clark & Word, 1974; Kazdin & Bryan, 1971), providing some of the first evidence of the influence of ability as a general concept. More recent research has identified self-efficacy as crucial in predicting prosocial behavior (Schwarzer et al., 1992), particularly knowing how to help (Gini et al., 2008; Pöyhönen et al., 2010).

The influence of general self-efficacy may be particularly relevant during a flight attendant assault. The environment's unique characteristics and the particulars of the triad may cause some to perceive the event as too specific for them to handle. Even passengers with special training or abilities may view the details of the environment and conflict as not matching the events they experienced in training. They may over-specify the event as "a flight attendant assault on an inflight aircraft full of passengers." However, those with high *general* self-efficacy may perceive the event in more general terms, such as "an assault "or "a woman in trouble." The generalization may better align with a passenger's general self-efficacy and result in greater confidence that they know how to respond. However, the generalization may not guarantee intervention.

Not all research has found self-efficacy to be influential. Some research has specifically found self-efficacy to be *not* correlated with helping behavior (Andreou & Metallidou, 2004; Rigby & Johnson, 2006; see Ashton & Severy, 1976) or correlated only in specific circumstances (Jouriles et al., 2016; Pöyhönen et al., 2010). These

conflicting results may indicate that the influence of self-efficacy fluctuates based on the scenario, familiarity with surroundings, the people involved, or other variables.

Perceived Social Influence

Social groups strongly influence how their members behave, albeit informally, through expectations. Among children, peer influence is a powerful force due to the close social ties between friends and the need for acceptance by the social group (Bukowski et al., 1996; Caravita et al., 2009; Cillessen & Mayeux, 2004; Goossens et al., 2006; Juvonen & Galvan 2008; Salmivalli et al., 1996). The pressure to follow behavioral expectations can also come from close family members (DeSmet et al., 2016). However, the influence is not exclusively toward prosocial actions. Social influence can equally manifest itself as either helpful or harmful behavior (Caravita et al., 2009; Cillessen & Rose, 2005; Espelage et al., 2003; Gini, 2006, 2007; Goossens et al., 2006; Juvonen & Galvan, 2008; Lease et al., 2002; Newcomb et al., 1993; Nickerson & Mele-Taylor, 2014; Parkhurst & Hopmeyer, 1998; Salmivalli et al., 1996; Sandstrom & Cillessen, 2006; Slee, 1994). The influence can even overwhelm personal inclinations toward remaining passive.

Lacking any feeling of personal responsibility to intervene can be overpowered by the expectations of close friends and family (Pozzoli & Gini, 2010). The pressure extends the boundary of responsibility to include not just those dictated by morals or ethics but also upholding behavioral expectations of close friends and family. The responsibility to intervene is not directed solely toward the victim but also toward satisfying expectations. Such a concept is uniquely tied to group dynamics because the social group need not be present to exert influence.

Merely perceiving what close friends and family expect can generate strong influences to act accordingly (Keefe, 1994; Rigby & Johnson, 2006; see Hardy & Carlo, 2005; see Pozzoli & Gini, 2010), creating a contrast to the diffusion of responsibility theory. Instead of responsibility being dispersed from a single person to all present, it involves a concentration of responsibility from a dislocated group onto a single person. During a flight attendant assault, a passenger's hesitancy to intervene may be overpowered by the perceived expectations of those they are traveling to visit. Even if a passenger-bystander feels no personal responsibility to intervene, they may feel a responsibility to adhere to the helping expectations of close friends and family. Of course, the opposite may be true as well. A passenger-bystander may feel some level of personal responsibility to help but may refrain from doing so to satisfy the expectations of friends and family to remain uninvolved.

The Complex Gender Variable

The inclusion of gender presents researchers with a unique dilemma. A significant amount of research has found gender to have only a marginal (Karakashian et al., 2006; Laner et al., 2001) or no effect (Benson et al., 1976; Darley & Latané, 1968; Fischer et al., 2011; Gruder & Cook, 1971; Howard & Crano, 1974; Macháčková et al., 2013; Morgan, 1978; Orlando, 2020; Rabow et al., 1990; Wiesenthal et al., 1983) on defending behavior. However, there is an equal amount of research showing otherwise. Significant discrepancies in first-order gender analyses make such findings useful only in establishing a baseline to understand higher-level correlations. The different results at these higher orders reveal that findings become relevant only to a particular scenario in a specific location with narrowly specified conditions. Indeed, understanding the role of

gender in helping is valuable, but any new bystander scenario should begin by evaluating gender differences at the most basic level.

The Prosocial Male

There is a significant amount of research showing the greater likelihood of males, in general, to exhibit helping behavior (Brewster & Tucker, 2016; Eagly & Crowley, 1986; Feinman, 1978; George et al., 1998; Hoffman, 1977; Laner et al., 2001; Piliavin et al., 1969, 1975). The likelihood is even more pronounced in very serious or dangerous situations (Austin, 1979; Belansky & Boggiano, 1994; Eagley & Crowley, 1986; Liebst et al., 2019; Shotland & Heinold, 1985). The cockpit voice recorder of Flight 93 captured exclusively male voices engaging in the counterattack against the hijackers (Hibbitts, 2006). Also, the only one to offer assistance to Genovese as her assailant attacked her was a male, albeit his help was only verbal and from a distance (Krajicek, 2011). Researchers have primarily relied on sociology to understand such findings.

Explanations for the above results have been speculative, identifying social expectations of heroics and chivalry as the most likely cause (see Eagly & Crowley, 1986). Such reasoning presents a unique variation to the self-centered bystander topic. In such cases, helping would not be driven by a motivation to reduce one's physiological distress but instead to avoid ridicule from others. It would mean the presence of others may not only act to diffuse responsibility but also aggregate and intensify unwanted moral judgment. Helping occurs when the perceived threat of social disapproval of passivity becomes too great. In non-physical situations, males generally help female victims the most (Eagly & Crowley, 1986; Feinman, 1978; Laner et al., 2001), which certainly supports the heroics and chivalry theory. However, they are also *less* likely to

intervene in a male-on-female assault than in other physical conflicts (Shotland & Straw, 1976).

The latter results seem remarkably atypical since male aggression toward females is predominantly viewed as reprehensible (Bethke & DeJoy, 1993). Passivity in a male-on-female assault may reveal a social expectation of males to refrain from publicly challenging another male's dominance (see Carlson, 2008). However, males often avoid intervening on behalf of other males as well. In this case, passivity toward a male victim may reveal an equally strong social expectation to avoid threatening another male's sense of masculinity or competency to handle the situation themselves (see Gusfield et al., 1981). The results, in aggregate, may indicate that male bystanders face competing normative social pressures influenced by the combination of situational characteristics and genders of the other two actors in the triad.

The dilemma makes the intervening decision for male bystanders highly complex. A myriad of different variables characterizes every situation. A male bystander must assign relative values to each variable to infer social expectations of intervention while predicting and evaluating the favorable and unfavorable judgment levels they will incur. The cognitive process may indicate that males are more attuned to the social pressure of other bystanders. Research directly relating to this possibility is peripheral at best, which makes accurately predicting male intervention behavior challenging.

The Prosocial Female

Notwithstanding the evidence of male propensity to intervene, there is also significant evidence to the contrary—that female bystanders are the ones more likely to help (Austin, 1979; Banyard, 2008; Belansky & Boggiano, 1994; Bihm et al., 1979;

Caravita et al., 2009; Gini et al., 2007; Ma, 2002; Monks et al., 2002; Nickerson et al., 2008; Oh & Hazler, 2009; Olweus, 1993; Piliavin et al., 1969; Poyhonen et al., 2010; Pozzoli & Gini, 2010; Rigby & Johnson, 2006; Salmivalli et al., 1996; Van Cleemput et al., 2014). Whereas dangerous situations elicit more male intervention, removing the danger makes females more likely to help (Banyard, 2007; Eagley & Crowley, 1986; see George et al., 1998). The reason may be a tendency for females to be more risk-averse. More specifically, females, more than males, may perceive themselves as unable to intervene in a physical attack without suffering harm. In an early bystander study involving a person demanding entry to a room, females responded much slower when alone than in a group (Levy et al., 1972). Researchers did not attempt to explain the apparent reverse bystander effect—ignoring the real possibility that a female alone in a room might feel highly threatened by a strange male voice demanding entry. Females also show seemingly contrary tendencies in other situations.

Generally, a known relationship between attacker and victim (husband-wife, parent-child, dog-owner) attenuates bystander intervention (Laner et al., 2001). For males, the hesitancy is uniform across all types of relationships, but females show an opposite tendency for the husband-wife pairing (Laner et al., 2001). Females are *more* likely to intervene in a husband-wife conflict than when a male stranger attacks a woman (Laner et al., 2001). Researchers surmise that the reversal is due to a fear among females that a stranger-attacker is more apt to transfer their aggression onto a female helper, while a husband or boyfriend is less likely to do so (Laner et al., 2001). The evidence supports the possibility that females, more than males, tend to be negatively influenced by the prospect of incurring harm. Alternatively, it may be that males and females are equally

affected by the threat of injury, but since the social expectation of heroics is absent for females, it does not overshadow such fears. Other social expectations do apply exclusively to females.

In cases where children are in distress, women are notably more likely to intervene (see Laner et al., 2001). The results may indicate a biological predisposition toward nurturing behavior (see Eagly & Crowley 1986), especially if victims are powerless. However, the results do not align with the inaction of dozens of women who ignored Bulger in his final hours. It may be that the tendency toward nurturing is not innate but instead a response to behavioral norms expecting such behavior from females more than males. The inaction of so many other women around Bulger may have convinced each one that the expectation of caring for the child was absent. In alternate situations, females may feel an expectation and a unique obligation to intervene.

Females are more likely than males to intervene in female-on-female assaults, and this inclination is particularly pronounced when other male bystanders are present (Lowe et al., 2012). Researchers found that in the presence of males, females feel a particular social obligation to stop the objectification and trivialization of women that occurs when men watch women fighting (Levine et al., 2020). These findings are novel because the pressure to intervene is directed toward satisfying a perceived obligation to a vast social class. Intervention occurs out of loyalty to a class of individuals, most of whom the bystander will never meet. During a female assault on a female flight attendant, the influence may result in high intervention rates for female passengers, mainly if they characterize the victim primarily as a woman rather than a flight attendant.

Higher-Order Effects

The conflicting evidence regarding male and female intervention rates by themselves has necessitated examining higher-order effects of gender. However, the complexities of gender intervention beyond first-order analysis make it difficult to identify patterns. Researchers have realized that many first-order correlations change by introducing other variables.

As victims, females are more likely than males to be helped by strangers (Benson et al., 1976; Eagly & Crowley, 1986; Howard & Crano, 1974; Laner et al., 2001; Shotland & Huston, 1979; Wiesenthal et al., 1983). However, when paired with the bystander, results also show that defenders are more likely to help those of the same gender (Saino et al., 2011). Even early bystander researchers found the impact of the victim's gender had unique two and three-way interactions with other variables (attacker's gender and perceived harm to victim) (Austin, 1979; Howard & Crano, 1974). Such evidence diminishes the value of first-order effects of gender since they become relevant only within the collection of controlled variables. Limited research has also found that the gender of other bystanders can influence how a person responds.

A review of CCTV footage revealed that bystanders are more likely to intervene when males are present during violent events, even if the other male bystanders are passive (Liebst, 2019). That dynamic may show that interveners have a greater expectation of receiving additional help from males (Fischer et al., 2011), but it is also possible that the reasons vary for different bystanders.

Females who intervene in the presence of males may perceive a reduced possibility of incurring harm. Such an inclination would exist if the female expects male

bystanders to come to her aid if the attacker turns their aggression toward her. For males, the presence of other males may create a perceived but unspoken challenge to their heroic abilities. Additional research using CCTV in a controlled setting also found the relationship between bystanders' genders complex.

After reviewing CCTV footage of violent assaults, males indicated they were more likely to respond if females were present, while females said they were more likely to respond if males were present (Levine & Crowther, 2008). However, the results are difficult to generalize since the research design only included footage of a male attacking a woman. Whether results would be different in other scenarios has not been researched.

Notwithstanding the gap in mixed-gender bystander research, it becomes apparent that the full effect of gender is a highly complex mix involving the genders of the victim, attacker, bystander, and co-bystander, which combine with the situation and context to create potentially sixth-order effects. However, relevant research past the second order is scant, possibly due to its complexity. Given the discrepancies of gender influence, researchers presenting novel situations should initially approach gender at the most basic level and expand only after discovering relevant differences.

Bystander Intervention Model Steps

A unique byproduct of gender specificity within the BIM steps is that it may explain many differences in bystander behavior. If one gender can better accomplish a particular step, they will be more likely to intervene.

Noticing the Event. Research tends to show females notice events more and with greater clarity than males. Adolescent girls notice bullying events at a significantly higher rate than boys (Jenkins & Nickerson, 2017), although researchers are not convinced of

the reasons. Adult females also have higher rates of recognizing child abuse (Hoefnagels & Zwikker, 2001), but explanations are likewise absent. The consistency across different ages and contexts may indicate either females are better at recognizing victimization characteristics or they focus more on the people in an event.

Recognizing the Need for Intervention. Females also show a higher recognition of a victim's need for help. In alignment with the previous step, adolescent girls are significantly better at interpreting a victim's needs during bullying incidents (Jenkins & Nickerson, 2017), and women clearly recognize a child abuse victim's need for help better than men (Hoefnagels & Zwikker, 2001). A female's ability to better interpret nonverbal cues and emotions (Hall, 1990) may make them particularly adept at accomplishing this step when there is no overt declaration of need from the victim. Adolescent boys who witness bullying have difficulty discerning the victim's needs (see Bastiaenses et al., 2014). However, the results are the opposite after introducing a bystander's history of victimization.

For adolescent girls, past victimization reduces their ability to recognize an event as an emergency (Jenkins & Nickerson, 2017). For adolescent boys, however, the opposite is true. Past victimization makes them more likely to correctly identify bullying emergencies (Jenkins & Nickerson, 2017). The difference may be due to how young boys and girls socialize and internalize victimization. Girls may be more apt to share their experiences with friends, making the memory of the event less threatening and the after-effects shorter-lasting. Boys, however, may tend to hide the incident for fear of being stigmatized as unable to defend themselves. The avoidance may create longer-lasting memories of anxiety (Carney et al., 2010; Nishina & Juvonen, 2005) and fear, making

boys more alert to future events. Admittedly, these results only apply to the narrow case of adolescent bullying, and the offered explanation is unresearched. However, the results, though limited, add to the complexity of the gender variable.

Taking Responsibility. A significant research gap exists regarding any relationship between gender and taking responsibility when faced with a helping event. The evidence suggests that the influence of adolescent peers *not* to help is much more substantial in boys than girls (Prinstein & Dodge 2008), though this tends to wane in adulthood. Males do appear to use victim-worthiness in their evaluation of responsibility more than females (Burn, 2009), but beyond this somewhat anecdotal evidence, there is scant research available to draw conclusions.

Knowing What to do. Much research into the relationship between bystander gender and knowing how to respond shows opposite tendencies. Generally, females of all ages show greater self-efficacy for defending (see George et al., 1998; see Gini et al., 2008). Beyond that, however, second-order effects of gender diverge.

Among adolescents, girls who had intervened in past bullying reported greater knowledge of how to do so, but the exact opposite was true for boys (Jenkins & Nickerson, 2017). Those boys who had intervened prior were less likely to report knowing how to respond (Jenkins & Nickerson, 2017). The reasons are speculative but potentially involve the more significant social pressure on adolescent boys to ignore bullying, which makes them the focus of more intervention education (Jenkins & Nickerson, 2017). The relationship between self-defense training and location is also the opposite for males than females.

College-aged males with self-defense training intervened more on campus than off, but females with self-defense training intervened more off-campus than on (Brewster & Tucker, 2016). Researchers did not provide possible explanations, but these and previous examples illustrate the highly diverse relationship gender has with other variables so much so that generalizations outside of the very narrow specifics of an event become questionable.

Deciding to Act. There is limited research addressing gender differences during this final decision to act. Researchers generally agree that females are more victim-centric in their attitudes (George et al., 1998; Gini et al., 2008; Hoffman, 1977; Rigby & Slee, 1991, 1993) and actions (Bastiaenses et al., 2014; Belansky & Boggiano, 1994; Brewin, 1984; Hastings et al., 2000; Hoffman, 1977; Van der Graaff et al., 2014) when confronted with the intervention decision. Although, when comparing levels of indirect support (deciding to tell authorities), the difference between males and females vanishes (Bastiaenses et al., 2014). Social influences on the decision to act indicate that forces *against* the decision are stronger for adolescent boys than girls (Cowie, 2000). That does not imply that boys provide more negative peer pressure, but only that it influences boys more. Since much of the existing bystander research omits the previous four steps and only includes the final one, it becomes unclear if differences between males and females are genuinely related to this step or a previous one. The omission strategy creates a significant gap regarding gender differences within the model steps.

Gaps in Existing Research

The research fills gaps within two categories—full and contextual. Full research gaps are wholly absent in research, while contextual research gaps are those related to applying an existing model or problem in an original context.

Full Research Gaps

Existing bystander research involving interpersonal conflict exclusively focuses on victims in a socially lower status (homeless person, immigrant, child, etc.) or state of lower power (student, female, animal, etc.). There is no research involving a victim in a higher social status or higher power state (police officer, teacher, politician, flight attendant). Based on their position, flight attendants hold a higher power over passengers regarding legal authority and experience with the functions of the aircraft cabin. An assault on a flight attendant constitutes an assault on a higher power figure—a dynamic not found in any bystander research. In addition, one of the roles of a flight attendant is to be a caregiver or provider of safety and security while passengers act as the recipient. A passenger intervening to help a flight attendant necessitates a reversal of these normative roles. Such a reversal is also absent in bystander research. Additional gaps involve the operationalization of constructs.

Researchers exclusively presume out-group bias to be negative. They likely presuppose that victimization is a result of negatively biased attitudes. While there is ample research showing the influence negative out-group bias has on helping behavior (see Brewer, 1979; see Doise et al., 1972; see Gardham & Brown, 2001; see Levine et al., 2002, 2020; see Rabbie & Horwitz, 1969; see Stephenson et al., 1976; see Tajfel, 1982;

see Tajfel et al., 1971), there is none testing the influence of positive out-group bias. Researchers have similarly restricted the application of nonintervention costs in research.

Researchers exclusively apply nonintervention costs to the victim. Standard designs present bystanders with a situation in which their failure to intervene results in harm or further harm solely to the victim. Limited research scenarios involve events such as smoke in the next room, but they all include the option to flee instead of intervening. There is no research in which the bystander is confined within the event without an opportunity to withdraw. Such a scenario confronts the bystander with a situation in which *they* incur the cost of their passivity. That dynamic, however, is not found in existing bystander research.

Contextual Research Gaps

There is no bystander research premised on an aircraft or any location characterized by the unique qualities of the aircraft cabin (confinement, social density, behavior requirements, noise level). These qualities place passenger bystanders in an environment not found elsewhere. In addition, the application of the BIM to varied scenarios is limited, and none involve a physical assault, an aircraft cabin, or any of the above characteristics. Although gender is an often-included variable in bystander research, the significant disagreements and confounding results make its impact widely misunderstood.

Existing research shows gender as a highly ambiguous yet salient variable. Results show significant variance at first and higher orders, resulting in a research gap regarding its influence. Research combining gender with BIM steps is minimal and encompasses only adolescent bullying, sexual harassment and assault, intoxicated

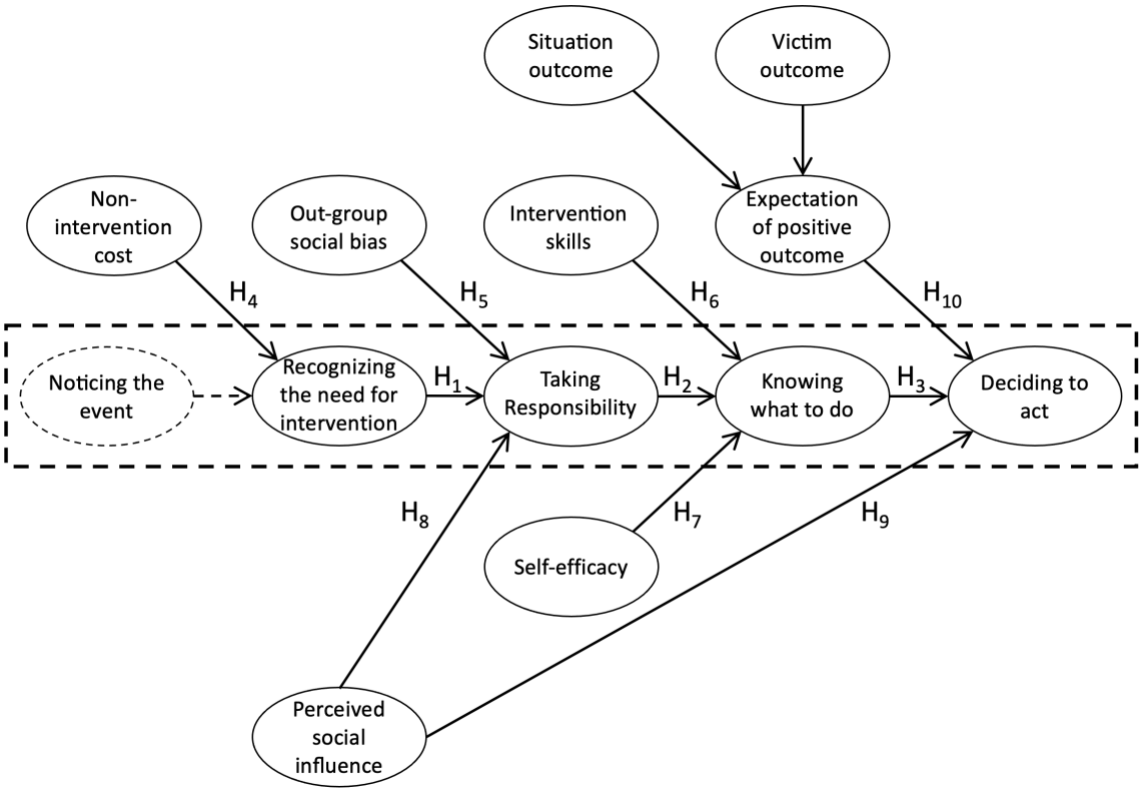
driving, and global refugee crises (see Albayrak-Aydemir & Gleibs, 2021; see Burn, 2009; see Jenkins & Nickerson, 2017; see Rabow, 1990). The presentation of a novel victim, location, environment, bystander experience, and unique operationalization of constructs necessitated a basic level analysis of gender to determine if any differences even existed.

Research Model and Hypotheses

The hypotheses were derived from similar research into bystander intervention. The first three relate to whether the bystander intervention model steps represent a bystander's cognitive process in a flight attendant assault scenario. The remaining hypotheses address whether factors found relevant in other bystander research also influence a flight attendant assault scenario. Figure 3 shows the full research model with associated hypothetical relationships.

Figure 3

The Research Model With Associated Hypothetical Relationships



Note. The dashed box represents the five steps of bystander intervention (Latané & Darley, 1970). While part of the five steps, *noticing the event* was excluded from the research, as the data-collection method prevented participants from failing this step. The step is included for clarity's sake only. *Expectation of positive outcome* is a hierarchical construct.

There is significant ambiguity regarding the influence of gender, its relationship within higher orders, and its varied relevance within different contexts—ranging from no effect to significant effect. The research established whether any difference exists between males and females for each hypothesis. To more clearly report results, each hypothesis has an associated sub-hypothesis postulating a difference between males and females.

Hypotheses Within the Bystander Intervention Model

Research into the steps of the BIM is limited, but for the most part, agrees that each step of the model influences the following action (see Albayrak-Aydemir & Gleibs, 2021; see Anker & Feeley, 2011; see Christy & Voigt, 1994; see Nickerson et al., 2014; see Rabow et al., 1990). However, not all relationships are conclusive (see Albayrak-Aydemir & Gleibs, 2021). Addressed scenarios include global emergencies; sexual harassment and assault; adolescent bullying; intoxicated driving; organ donation; political support; environmental consciousness; and child abuse (see Albayrak-Aydemir & Gleibs, 2021; see Anker & Feeley, 2011; see Burn, 2009; see Christy & Voigt, 1994; see Nickerson et al., 2014; see Rabow et al., 1990). Although the full model includes all five steps, the research excluded the first step of the model (noticing the event) since the data collection method did not allow participants to fail this step.

Existing research establishes the positive unidirectional relationship between steps two (recognizing the need for intervention) and three (taking responsibility) of the BIM in several situations (see Albayrak-Aydemir & Gleibs, 2021; see Nickerson et al., 2014; see Rabow et al., 1990; see Anker & Feeley, 2011). In this study, a passenger-bystander recognizing that a flight attendant needs help was hypothesized to influence them to feel more responsible.

H_1. Recognizing the need for intervention significantly positively influences taking responsibility.

$H_{1\text{-}1}$. There is a significant difference between males and females in the relationship between recognizing the need for intervention and taking responsibility.

Similar to H_1, several researchers have identified a positive, one-way relationship between steps three (taking responsibility) and four (knowing what to do) of the BIM (see Albayrak-Aydemir & Gleibs, 2021; see Nickerson et al., 2014; see Rabow et al., 1990; see Anker & Feeley, 2011). In this study, a passenger bystander taking responsibility to intervene was hypothesized to make them more likely to believe they know how to respond.

H_2. Taking responsibility significantly positively influences knowing what to do.

$H_{2\text{-}1}$. There is a significant difference between males and females in the relationship between taking responsibility and knowing what to do.

Hypothesis three is unique because it is the only BIM relationship with conflicting results for different scenarios. Researchers identified the predictive relationship in bullying, sexual harassment, intoxicated driving, and child abuse (see Christy & Voigt, 1994; see Nickerson et al., 2014; see Rabow et al., 1990), but specifically *not* in the context of global refugee intervention (see Albayrak-Aydemir & Gleibs, 2021). A flight attendant assault more closely aligns with the former situations since the bystander would be an in-person witness of the victimization. Bystanders of a global refugee crisis are, for the most part, distant observers of the problem. In this research, a passenger knowing how to help a flight attendant during an assault was hypothesized to make them more likely to decide to do so.

H_3. Knowing what to do significantly positively influences deciding to act.

$H_{3\text{-}1}$. There is a significant difference between males and females in the relationship between knowing what to do and deciding to act.

Hypotheses of Exogenous Constructs

The remaining seven hypotheses also come from existing bystander research. A large amount of research identifies the positive influence of the constructs in their designs, but only as predictors of a single intention to intervene. A smaller number of studies specify the relationship to an individual step of the BIM.

Fischer et al.'s (2011) metadata research, as well as CCTV footage (Liebst, 2019), found that when the cost of nonintervention is high (as in violent encounters), the bystander effect fades. A high cost of nonintervention reduces any ambiguity surrounding the event's seriousness. However, the cost of nonintervention in existing research is exclusively bound to the victim. Anecdotal evidence (Flight 93) indicates that a high cost of nonintervention incurred by a bystander will have a similar impact on reducing ambiguity surrounding whether an event necessitates intervention. The cost of nonintervention during a flight attendant assault was hypothesized to positively influence a passenger's interpretation of the event needing intervention.

H_4. Nonintervention cost significantly positively influences recognizing the need for intervention.

$H_{4\text{-}1}$. There is a significant difference between males and females in the relationship between nonintervention cost and recognizing the need for intervention.

People are more inclined to accept responsibility for helping those who are part of their social group (Burn, 2009; Levine et al., 2002) or those who are similar to themselves (Albayrak-Aydemir & Gleibs, 2021; Emswiller et al., 1971). The paradigm of in-group bias being synonymous with favorability and out-group bias with negativity

has created a significant gap regarding their application in measuring how victims are perceived. In nearly all bystander research, the victim occupies a lower status or power position. That pattern advances the assumption that bias toward victims must be both derogatory and due to their social standing. A more accurate characterization of bias should consider the socially-attributed traits of the out-group members, which may not always be negative.

Passengers expect flight attendants to be friendly, hard-working, knowledgeable, and well-groomed (see Tsaur et al., 2020). Since passengers expect these traits of the entire group, it can create a favorable bias toward how individual flight attendants are perceived. The favorable bias toward flight attendants was hypothesized to positively influence a person to take responsibility for intervening.

H_5. Out-group social bias significantly positively influences taking responsibility.

$H_{5\text{-}1}$. There is a significant difference between males and females in the relationship between out-group social bias and taking responsibility.

A lack of acquired intervention skills can hinder a bystander's ability to formulate an intervention plan (Albayrak-Aydemir & Gleibs, 2021; Burn, 2009; Cramer et al., 1988). In sexual assault incidents, skills deficiency is one of the most significant barriers to knowing how to respond (Burn, 2009). In that context, the influence of intervention skills was unaffected by gender (Burn, 2009). In this study, having a high level of intervention skills was hypothesized to positively influence a passenger’s belief that they know what needs to be done to help a flight attendant under attack.

H_6. Intervention skills significantly positively influences knowing what to do.

$H_{6\text{-}1}$. There is a significant difference between males and females in the relationship between intervention skills and knowing what to do.

Researchers will often measure self-efficacy specific to the event under study (see DeSmet et al., 2016). Pöyhönen et al.'s (2012) bullying research identified a positive relationship between defending behavior and a belief in one's ability to stop bullying. Applying this specificity is problematic since the characterization of an event rests with the observer, not necessarily the researcher. A researcher may characterize an event as bullying, but a bystander may perceive it otherwise (teasing or felonious assault). Applying a bullying efficacy scale to either perception reduces validity and increases conclusion errors. In other bullying research, Gini et al.'s (2008) solution was to utilize a social self-efficacy scale measuring a belief in one's ability to be successful in broad social situations (getting people to listen and expressing opinions in public). Their results showed a positive relationship between social self-efficacy and defending behavior. However, their use of social self-efficacy still provided a specification of bullying as a social issue, which may not be how all bystanders classify the event.

The broadest measure of self-efficacy is general self-efficacy. It measures a person's belief in their general ability to achieve desired results effectively (Bandura, 1997). Using this generalized measure minimizes differences between researcher and participant characterization of an event. In this research, general self-efficacy was hypothesized to positively influence knowing what to do. The distinction is that *knowing what to do* refers specifically to the described assault on the flight attendant. A person who believes that they are generally able to accomplish their goals is more likely to be able to develop a specific plan of intervention.

H_7. Self-efficacy significantly positively influences knowing what to do.

$H_{7\text{-}1}$. There is a significant difference between males and females in the relationship between self-efficacy and knowing what to do.

Those individuals with whom a person has close social ties influence their attitudes toward helping (see Keefe, 1994). Even if a person feels no personal responsibility to help, the perceived pressure from close friends and family can influence them to do so (see Pozzoli & Gini, 2010). Researchers commonly include friends and parents as those most influential to a person (see DeSmet et al., 2016; see Keefe, 1994; see Pozzoli & Gini, 2013; see Rigby & Johnson, 2006). Researchers also have found the expectations of friends and parents to be significant influencers of the level of responsibility a person feels to defend a victim (see Pozzoli & Gini, 2010, 2013). However, these findings are not universal (see DeSmet et al., 2016). The perceived social influence from friends and parents in this research was hypothesized to influence a participant to feel more personally responsible for assisting a flight attendant during an assault.

H_8. Perceived social influence significantly positively influences taking responsibility.

$H_{8\text{-}1}$. There is a significant difference between males and females in the relationship between social influence and taking responsibility.

Similar to H_8, those with whom a person has close social ties influences their behavior (see Cillessen & Mayeux, 2004). Researchers have found those expectations to be positively linked to active prosocial intervention (Pozzoli & Gini, 2010, 2013; Rigby & Johnson, 2006). In this research, the perceived influence from friends and parents was

hypothesized to positively influence a participant to decide to intervene during a flight attendant during an assault.

H_9. Perceived social influence significantly positively influences deciding to act.

$H_{9\text{-}1}$. There is a significant difference between males and females in the relationship between social influence and deciding to act.

If bystanders expect their intervention to result in a positive outcome, they are more likely to intervene. Research has shown this expectation centers around making the victim feel better (see Gini, 2006; see Pöyhönen et al., 2012; see DeSmet et al., 2016) and stopping the harmful event (see Gini, 2006; see Pöyhönen et al., 2012). Notably, Desmet et al. (2016) found the victim-centered expectation to be the strongest predictor of intervention behavior. In this research, a bystander who feels their intervention will result in an improved outcome was hypothesized to be more likely to decide to take action.

H_{10}. Expectation of positive outcome significantly positively influences deciding to act.

$H_{10\text{-}1}$. There is a significant difference between males and females in the relationship between expectation of positive outcome and deciding to act.

Summary

Chapter II explained several aspects of the underlying theoretical foundation, the model of bystander intervention, research constructs, the role of gender in a bystander event, gaps in existing research, and the hypotheses. The chapter used several historical examples to illustrate first how observable characteristics of the assailant, victim, and bystander may affect intervention. Next, analyzing the relational ties between event actors revealed the nuances of human interaction during helping events. The chapter also

examined the complex situational intricacies of an event and showed how the various items might manifest themselves during a flight attendant assault. The review included recent challenges to bystander theory due to more modern data collection methods. The first half closed with a presentation and associated explanation of the basic bystander intervention model.

The chapter relied on historical examples to explain how failures of the BIM steps might occur. It also reviewed the research constructs and explained how they relate to bystander behavior. An extensive review of the intricacies of gender in bystander research followed, including its possible influence during each step of the BIM. The chapter highlighted several research gaps, some entirely absent in research, while others encompass original applications of the BIM or constructs. The section closed by explaining the hypotheses and their origins.

Chapter III: Methodology

Chapter III reviews the details of the research method, techniques used to identify the necessary amount of data, and specific procedures used to collect it within ethical guidelines. It also describes the measurement instrument and the plan followed to evaluate its reliability, validity, and invariance. There is a description of how heterogeneity was identified and managed and an explanation of the multigroup analysis. The chapter closes with details of how the explanatory and predictive power of the structural model were assessed.

Research Method

The research used a quantitative method. Quantitative methods are those which use numerical data to evaluate relationships between variables. These relationships are often based on existing theories or models and are characterized as correlational, causal, or influential. When based on scientific literature, the objective of quantitative research is often to expand a theory or model's application by introducing new variables, relationships, or contexts.

A significant amount of quantitative bystander research entails contexts similar to the present study. These existing studies focused on a specific population, such as adolescents (see Bastiaenses et al., 2014), college students (see Wiesenthal et al., 1983), or subway riders (see Greitemeyer et al., 2006), and incorporated specific variables such as efficacy (see Banyard, 2008), group-bias (see Abbott & Cameron, 2014), and helping skills (see Huston et al., 1981). Researchers examined these variables within a specific context, such as bullying (see Dillon & Bushman, 2015), sexual assault (see Kleinsasser et al., 2015), and refugee aid (see Albayrak-Aydemir & Gleibs, 2021). Each of these

studies used a quantitative research method. While many populations, variables, and contexts were familiar, their analyses generated unique results. The data in this research is also numerical and involves identifying influential relationships between variables. These details made a quantitative method the appropriate choice to add to the body of existing quantitative bystander literature.

Research Design

The research used a non-experimental research design. Such designs do not meet the requirements of experimental (independent groups, random group assignment, and an applied treatment) or quasi-experimental (containing two experimental conditions) research. Although this research does include a multi-group analysis between independent groups (males and females), there was not random assignment to those groups or any application of group-unique treatment. The extent of the analysis was merely to determine if group differences existed.

Analysis Approach

The research used a structural equation modeling (SEM) approach. SEM is a type of multivariate analysis unique in its ability to visually model relationships between several observed and unobserved (latent) variables (Byrne, 2016). It is particularly beneficial for explaining relationships between variables that may be simultaneously dependent and independent (Hair et al., 2010). For example, out-group social bias may influence *taking responsibility* (BIM step three), which in turn may influence *knowing what to do* (BIM step four), which then may influence *deciding to act* (BIM step five). BIM steps three and four simultaneously have dependent and independent relationships

with other variables. SEM also includes various subset characteristics describing its theoretical approach and analytic strategy.

Confirmatory vs. Exploratory Factor Analysis

Within the SEM framework, researchers analyze outer models (measurement models) using either a confirmatory or exploratory factor analysis approach. Confirmatory factor analysis (CFA) is a theory-driven approach to SEM in which researchers primarily use a priori relationships to specify constructs and models (Hair et al., 2022). Exploratory factor analysis (EFA), on the other hand, is a data-driven approach used when there is no presupposed pattern in the data or construct relationships (Hair et al., 2022), and the goal of the research is to explore such data for new patterns (Hair et al., 2010).

During CFA, researchers use existing literature to specify the structure of the inner and outer models before computing results (Hair et al., 2010). However, during EFA, the relationships within observed and latent variables are established based on the data with no presumption. Since this research used measurement and structural models from literature, it was a confirmatory analysis. Although relationships between observed and latent variables were pre-established through prior research, the researcher confirmed them during the pilot studies.

Difference Between Covariance-Based and Partial Least Squares-Structural Equation Modeling

The two primary strategies for model estimation are covariance-based (CB) and partial-least-squares-based (PLS) methods. The core difference between CB-SEM and PLS-SEM is how the associated algorithm estimates the constructs. The CB algorithm

considers all latent variables as factors that influence, or explain, the variance of their indicators. During model estimation, the CB algorithm restricts its consideration to only the shared variance within each measurement model. Thus, for estimations to be valid, indicators must have a high level of shared variance. The characteristic means that while, in theory, a CB-SEM model can include formative constructs, results will be misleading since indicators of a formative construct need not (and should not) covary (Hair et al., 2022).

The PLS algorithm, however, considers all variance during estimation. In other words, PLS attempts to explain as much indicator variance as possible, not just the shared portion. That detail allows for formative constructs since the PLS algorithm considers more indicator variance than just the shared portion. A byproduct of this full-variance approach is that PLS models tend to achieve better out-of-sample predictions than CB models, though the difference appears to narrow, or even equalize, with large sample sizes (see Reinartz et al., 2009).

Research Procedures

The research followed an eight-step process. The process began after establishing research questions, hypotheses, population, and model specification. Steps four and five were repeated based on pilot study results.

1. Develop measurement instrument
2. Identify sampling frame
3. Obtain Institutional Review Board (IRB) approval
4. Validate the instrument (pilot study)
5. Modify the instrument (if required)

6. Collect data
7. Analyze data (including multi-group analysis)
8. Report results

Population and Sample

A research population is a group to which results are generalized. Only a relatively small number of people in a population are used as participants, resulting in a sample. The method to identify sample participants defines a sampling frame, and contacting each participant is a sampling strategy.

Population

The target population was American adult (at least 18 years old) airline passengers. The exclusion of non-adult participants does not imply that an inherent shift in perception occurs at 18 but was only stipulated to simplify the consent process. The restriction on nationality was an effort to minimize the compositional variance of the constructs due to cultural differences.

The term *airline passenger* means someone who travels frequently enough to be familiar with the inflight environment. The frequency required to maintain familiarity is, of course, highly subjective. A person whose most recent travel was many years ago may not remember the details enough to be able to reimagine the experience. To establish a measure of frequency required to be considered an airline passenger, the researcher reviewed the customer reward programs of 28 global airlines. Most contained a necessary frequency of travel to avoid the expiration of rewards (Kunesh, 2022). While it is unlikely airlines based these times on any psychological basis of experience-recall, it does indicate what those in the industry consider to be a consistent traveler. Excluding

those airlines whose rewards had expiration dates of 10 years or more resulted in an average expiration of 2.4 years (n = 20). Hence, participants were limited to those who had traveled within the previous two years.

Sample Size

Selecting a sample size involves a balance between statistical strength and practicality. Researchers attempt to identify the minimum sample size that allows their statistical conclusions to have minimal error while also avoiding expending unnecessary resources. In general, PLS methods require smaller samples for model convergence (compared to CB methods), but researchers should be cautious about using PLS purely for that reason.

Since PLS treats constructs as composite factors (as opposed to common factors), researchers can achieve successful convergence with sample sizes smaller than those required by CB methods. However, model convergence does not indicate generalizability or the absence of type I or II errors. Researchers should recognize that small sample sizes in and of themselves should neither be the goal nor accepted purely for convenience. However, determining an appropriate sample size, particularly for SEM, can be challenging due to the lack of a universally-accepted method.

There are several techniques for determining the minimum sample size. Some emphasize simplicity, while others provide a more complex statistical basis. Within the latter, some techniques emphasize the number of variables, while others emphasize the relationship between them. Researchers should select the method that best supports their requirements.

Rule of 10 Method. The rule of 10 is the simplest method and is based only on the most complex multiple regression in either the measurement or structural model (Barclay et al., 1995). The rule dictates a minimum of 10 observations for each predictor of either a formative or endogenous construct—whichever one has the most (Barclay et al., 1995). The technique is relationship-based (i.e., relying on the number of relationships), considering each construct's predictors independently from the overall model structure. Applying the *rule of 10* to the present model would have resulted in a minimum sample size of $n = 50$ (five indicators each for the formative constructs *out-group social bias* and *intervention skills*). Although the technique is easy to use, it lacks a statistical basis, even as initially proposed (see Nunnally, 1967).

Minimum R^2 Method. A more robust method for minimum sample size is the *minimum* R^2 method. The method is primarily based on the significance level, statistical power, minimum expected R^2, and the most complex relationship structure of any latent variable (i.e., the maximum number of arrows pointing to a latent variable) (Kock & Hadaya, 2018). The method is variable-based, as its central focus is a statistical characteristic of the model's dependent (endogenous) variables. Although the method developers do not provide the formula, they provide a table for reference (see Kock & Hadaya, 2018).

Applying the minimum R^2 method to this research would have resulted in a minimum sample size of $n = 124$ ($R^2 = .1$, power = .8, $p = .05$, three relationships). The method is an improvement over the *rule of 10* as it incorporates a justification based on statistical significance, power, and a characteristic of the structural model. However, its

reliance on R^2 as the core discriminator may only partially align with models whose primary focus is the predictors (i.e., path coefficients) rather than their targets.

Inverse Square Root Method. The inverse square root method is similar to the *minimum* R^2 method's inclusion of statistical power and significance. However, it differs in its emphasis on the relationships between constructs (path coefficients) as the core discriminator rather than on dependent variables. That aspect makes it a relationship-based method as opposed to a variable-based one. It utilizes a ratio of path coefficient to sample size to derive a value that reduces type II error rate to the researcher-defined acceptable level (Kock & Hadaya, 2018). Using a statistical power of .8 and p = .05, the inverse square root formula is (Kock & Hadaya, 2018):

$$n = \left(\frac{2.486}{|\beta|}\right)^2 \quad (1)$$

Where:

β = the minimum anticipated path coefficient

Applying the *inverse square root* method to this research would have resulted in a minimum sample size of n = 241 (β = .16, power = .8, p = .05). The minimum anticipated path coefficient is based on Anker and Feeley's (2011) research into green living in which they identified a path coefficient of .16 between BIM steps *taking responsibility* and *knowing what to do*.

Compared to the *minimum* R^2 method, the inverse square root method is more applicable to models designed to identify the strength of specific predictor variables over models designed to minimize unobserved variance. However, identifying a path coefficient a priori requires making an assumption of results before data collection even begins (Hair et al., 2022). Absent this step, researchers can retroactively derive a

minimum detectible relationship based on the number of observations. However, doing so requires caution, as researchers can become biased toward making type I errors. The method also makes no allowances for the number of observed variables per latent construct.

Ratio Method. Early researchers proposed that the relationship between the numbers of observed and latent variables rather than their independent quantities is more relevant in identifying minimum sample sizes (see Boomsma, 1982; see Marsh & Bailey, 1991). After reviewing several Monte Carlo simulations, Westland (2010) derived a ratio-method formula for the minimum sample size required to reduce the chance of committing a type II error to 5%:

$$n = 50\left(\frac{p}{k}\right)^2 - 450\left(\frac{p}{k}\right) + 1{,}100 \qquad (2)$$

Where:

p = the number of observed variables in the full model

k = the number of latent variables in the full model

The formula was not created from a deductive statistical basis but rather inductively from combining the results of multiple (35,000) Monte Carlo simulations (Westland, 2010). That fact makes the formula less a statistically-based rule than a guide, albeit justifiable via 35,000 repetitions.

The method optimizes at a ratio of $p / k = 4.5$ ($n = 88$), returning the smallest required sample size at this ratio. Both larger and smaller ratios require a greater number of samples, though Westland (2010) does not offer a practical explanation for the parabolic nature of the formula. Applying the ratio method to the model in this research

would have resulted in a minimum sample size of n = 88 (46 observed variables, 10 latent constructs).

The ratio method, however, considers all measurement models in aggregate by assuming the number of observed variables is equally distributed amongst latent constructs. It makes no allowance for differing numbers of indicators assigned to different latent variables. Such an omission may result in inaccurate sample size returns (and thus misleading results) for some portions of a model. For example, estimating the relationship between *nonintervention cost* and *recognizing the need for intervention* involves six observed variables and two latent constructs. Utilizing the ratio method for only this relationship would necessitate a sample size of n = 200, even though the aggregate model implies a minimum sample size of only n = 88. The model partition may converge with 88 samples but may lead to an erroneous conclusion (type I or II error).

Paired Latent Variables Method. Perhaps in response to the deficiencies of other techniques, Westland (2010) developed a minimum sample size algorithm based on the number of observed and latent variables, desired statistical significance and power, and the minimum effect size the researcher wishes to detect. The method has similarities to each of the previous four methods. However, it is unique in its inclusion of both measurement model specifications (numbers of indicators and latent constructs) and anticipated structural model estimations (path coefficients). This makes the *paired latent variables* method the only one that is equally variable and relationship-based. Applying the *paired latent variable*s method to the model in this research would have resulted in a minimum sample size of n = 765 (β = .16, power = .8, p = .05, 10 latent constructs, 46 observed variables). Like in the *inverse square root* method, the path coefficient (β) is

based on Anker and Feeley's (2011) research into green living, in which they identified a path coefficient of .16 between BIM steps *taking responsibility* and *knowing what to do*.

The *paired latent variable*s method, like the ratio method, suffers from the assumption that the number of observed variables is distributed equally across the entire model. Such an assumption can lead to erroneously high or low sample sizes for some relationships in the model. For example, in the present model, *recognizing the need for intervention* and its antecedent *nonintervention cost* includes two latent constructs and six observed variables. Keeping all other parameters constant returns a minimum sample size of $n = 359$. However, *knowing what to do* and its antecedents (*intervention skills*, *taking responsibility*, and *self-efficacy*) include four latent constructs and 21 observed variables. That model partition returns a minimum sample size of $n = 550$.

The parsing of the model into endogenous partitions creates a more accurate indication of the number of observations needed to detect the desired effect for the paths to each endogenous variable. It also aligns with the model estimation iterative process in which an endogenous construct's variance is regressed onto its predictors but not past them. The practice is more appropriate during CFA (as opposed to EFA) since the model specifies that many exogenous constructs have no relationship to many endogenous ones. Since these relationships do not exist, applying an assumption of equally distributed indicators throughout the entire model is inappropriate.

Considering all methods, the *paired latent variables* method includes the most input variables, is statistically based, and includes the anticipated path coefficient as a core discriminator. Although it also returns the largest sample size, it was selected for this research. An online sample size calculator (Soper, 2022) using the formula from

Westland (2010) returned a sample size of $n = 550$ for the largest model partition. Since multi-group analysis requires independent model estimation of each group, the researcher established $n = 550$ as the minimum sample size for each group. As a result, the aggregate minimum sample size was $n = 1{,}100$. To allow up to 5% of the data to be unusable, the researcher planned to collect a minimum of 1,158 observations. See Table C1 for a complete list of minimum sample sizes for all methods, the aggregate model, and all model partitions.

The researcher conducted a series of pilot studies to verify the reliability and validity of the instrument. Only the measurement models were evaluated during this phase, not the relationships between constructs. Using the same values for power, significance, and usable data with a minimum effect of $\beta = .1$ returned a minimum sample size of $n = 92$.

Sampling Frame

The sample frame was users of Amazon's® Mechanical Turk (MTurk)—an online crowdsourcing platform with approximately 500,000 participants. Due to the inclusion of a participant's perceptions of flight attendants (out-group social bias) as a variable, those who had worked in the airline industry (i.e., pilot, flight attendant, gate agent, crew scheduler) were excluded from the sampling frame. Such participants would have had unique interactions with flight attendants during employment. Those interactions would have been as coworkers—a markedly different relationship than that experienced by passengers.

Sampling Strategy

Stratified convenience sampling was utilized to identify an equal number of male and female participants while making no in-person contact. To minimize the number of unusable surveys (due to incompleteness or inattentive respondents), the researcher limited participants to those with an approval rate of 98% and a task completion history above 1,000. Current MTurk requesters identified a high approval rate and task completion history as two characteristics that significantly affect response quality (Amazon Mechanical Turk, 2019). Participants were compensated $1 upon survey completion.

Data Collection Process

Data collection occurred during two phases—the pilot studies and the main study. The researcher used the pilot studies to develop scales, verify the validity of constructs, and ensure the operability of the survey instrument. The focus of the main study was the estimation of the entire model. Data collection for both was conducted identically, though only the outer (measurement) models were evaluated during the pilot portions. To avoid biasing the analysis of the main study, model estimation did not occur during the pilot studies.

The collection process for both phases began with an informed consent statement followed by brief survey instructions. Participants were informed of the desire for all survey responses to be completed, though some demographic questions included a “skip” option without penalty. For both phases, the researcher passively solicited participants via stratified convenience sampling by posting a work request on MTurk. To ensure sufficient numbers of males and females for multi-group analysis, the researcher created

two identical surveys and restricted each to either only males or females. Having near-equal sample sizes ensured greater statistical power and more accurate model estimation (Hair et al., 2022). Stratification was via MTurk's qualification settings, allowing only participants who meet specific criteria to complete the task. However, participants still provided demographic data to verify eligibility and group membership.

Advantages of MTurk

MTurk allows affordable and efficient access to a large portion of the population (Rice et al., 2017; Sheehan, 2018). That characteristic affords a higher degree of generalization. At the participant level, the strategy of using MTurk allows for greater participant anonymity (Rice et al., 2017) while still assuring a high level of response quality (Casler et al., 2013; Farrell et al., 2017; Hauser & Schwarz, 2016; Rice et al., 2017). Researchers can strengthen this quality further by establishing minimum MTurk user scores (based on evaluative performance) required for participation. Results of MTurk's use as a data collection device show the demographic generalizability of samples to be high (Casler et al., 2013; Sheehan, 2018) while at the same time showing no significant difference in experimental results between MTurk and in-person participants (Casler et al., 2013). However, there are disadvantages associated with its use.

Disadvantages of MTurk

Using a specific commercial platform over a specific medium during a specific time frame can undoubtedly introduce a bias toward the type of person available for contact (Harms & DeSimone, 2015; Rice et al., 2017). The option to participate will not be available to the portions of the population that do not regularly use computers, the

internet, or MTurk. However, in terms of the population, individuals who do not use computers or the internet are also unlikely to be regular airline travelers. Gathering data via an online survey, however, does have unique disadvantages.

Using MTurk eliminates any ability to directly assess participant motivation or attentiveness (Rice et al., 2017). There is no guarantee that participants will not perform other distracting tasks in conjunction with accomplishing the survey. Research also shows that the employment characteristics of MTurk users are more information- and financial-industry-focused than the general population (Harms & DeSimone, 2015). That characteristic may result in a skewed representation of personality types.

Sources of Data

All data for both the pilot and main study were from the online survey. See Appendix B for a depiction of how the survey appeared to participants. To the extent that MTurk acted as an intermediary medium, the researcher directly collected all data.

Survey Procedure

The request for MTurk users to participate included a broad description of the task (a survey), the expected time for completion, and the reward for doing so. Users agreeing to participate and giving consent were presented with a four-section survey. The first section entailed collecting demographic data and establishing eligibility criteria. Eligibility questions ascertained that the participant was at least 18 years old, was an air traveler (had flown in the past two years), and had never worked in the commercial airline industry. If a participant failed to meet any eligibility requirement, they were thanked for their time and excluded from the remainder of the survey. The remaining

demographic information included annual income, gender, ethnicity, education level, and marital status.

For sections two and three, participants did not have the option to return to previous question blocks, and all questions were optional. Section two comprised the constructs *self-efficacy*, *intervention skills*, *perceived social influence*, *situation outcome*, and *victim outcome*. These constructs are independent of the flight attendant assault scenario in that they are the normal personality-based state of the participant, unaffected by any disturbance. To measure these participant attributes in an unaffected state, they were presented before the flight attendant assault scenario.

The questions for each construct were grouped but randomized. In other words, the participant answered all questions for one construct in random order before being presented with questions related to another construct. The order in which the constructs in this section were presented to participants was also randomized. The questions for *out-group social bias* were presented next. The purpose of presenting this construct after section two was that it narrowed the research's context and focus to that of flight attendants, and any subsequent questions were likely answered with flight attendants in mind.

For the third section, the participant was asked to imagine they were a passenger on an airline inflight. They were presented with the following written flight attendant assault scenario:

"You are sitting in an aisle seat on a mostly full airplane approximately halfway through your flight. You hear loud cursing and look up to see a passenger and a flight attendant, both women, standing in the aisle arguing a few rows away. You cannot tell

what the argument is about, but you see the passenger begin trying to punch the flight attendant as she continues to yell at her." The vignette remained on the screen for the remainder of the survey, allowing the participant to refer to it as needed.

Section three included, in order, *nonintervention cost*, *recognizing the need for intervention*, *taking responsibility*, *knowing what to do*, and *deciding to act*. However, the questions within each of these constructs was randomized. Presenting these constructs in order aligns with BIM theory that accomplishing each step is a prerequisite for its follow-on step. The underlying theory of the BIM is that a person will decide to act *only* after accomplishing the previous steps in order. Thus, it was more appropriate to present this section's constructs in the order in which a participant would cognitively perform them instead of randomly.

The final section was an optional open-ended response asking for impressions of the survey and assault vignette. The results of this question were only used to aid in potential follow-on research. The final screen included a short sentence thanking the participant for their time and attention.

Measurement Instrument

The measurement instrument was a 58-question survey (see Appendix B). It was designed using Qualtrics Core XM® and administered online. Qualtrics functionality allows for randomizing groups within groups (block randomization). The ability allowed a construct's indicators to be grouped and randomized while also randomizing the order of constructs. Qualtrics also allows skip-logic, relieving participants from having to continue if they fail to meet demographic requirements, although MTurk also integrates

this logic into its functionality. After data was collected, Qualtrics allowed for download in several formats (i.e., csv, sav, xls).

Constructs and Indicators

Each construct was measured via a set of observable indicators adapted from various literature-based bystander scenarios (see Table 2). Table C2 includes the original and modified phrasing of indicators.

Table 2

Model Constructs, Their Indicators, and the Context From Which They are Adapted

Construct	Indicators	Original Context
	Formative	
Out-group social bias	SB1, SB2, SB3, SB4, SB5, SB6[a]	Immigrant bullying
Intervention skills	IS1, IS2, IS3, IS4, IS5, IS6[a]	Violent crime
Nonintervention cost	NC1, NC2, NC3, NC4, NC5[a]	-[b]
	Reflective	
Perceived social influence	PS1, PS2, PS3, PS4, PS5	Adolescent bullying
Self-efficacy	SE1, SE2, SE3, SE4, SE5, SE6, SE7, SE8, SE9, SE10	Unmodified
Expectation of positive outcome	EO1[a]	
Situation outcome	SO1, SO2, SO3	Adolescent bullying
Victim outcome	VO1, VO2, VO3	Adolescent bullying
Recognizing the need for intervention	RN1, RN2, RN3	Refugee aid
Taking responsibility	TR1, TR2, TR3	Sexual harassment, adolescent bullying
Knowing what to do	KW1, KW2, KW3	Refugee aid
Deciding to act	DA1, DA2	Sexual assault

Note. See Table C2 for the original and modified indicator context.

[a] IS6, SB6, EO1, and NC5 are global indicators used only to evaluate the convergent validity of the instrument and were not included during model estimation.

[b] In literature, *nonintervention cost* is traditionally victim-centric. There is no equivalent bystander-centric measure in literature.

Scales

The research utilized 5-anchor Likert scales for all measurement models and considered the scales to be interval. The distinction is relevant due to statistical assumptions in the data analysis phase (Chyung et al., 2017). There is no guarantee, however, that respondents perceived the scale's anchors as having interval spacing. Some participants may have subconsciously assigned unequal differences between anchors, making the scale ordinal (Chyung et al., 2017). However, since there is no practical method to verify participant perception, the researcher assumed that participants treated the scale as interval.

Number of Anchors. There is significant ambiguity concerning the optimum number of anchors, or choices, in a Likert scale. Some researchers have found that reliability increases with more response options (Preston & Colman, 2000), peaking at a scale size of seven (McKelvie, 1978; Nunnally, 1967; Preston & Colman, 2000). However, the findings are not universally accepted, as some researchers assert that reliability and the number of response options are unrelated (Brown et al., 1991; Matell & Jacoby, 1971). Researchers also have differing views on the validity of various Likert scales. Some maintain that validity is unaffected by the number of choices (Matell & Jacoby, 1971). Others found higher validity coefficients in scales containing at least five anchors (Chang, 1994; Preston & Colman, 2000). The broad range of conclusions means that the type of respondent, context, medium, and other characteristics likely have unique impacts on an instrument's reliability and validity. Such a realization makes evaluating these criteria critical each time an instrument is used. Researchers should also leverage reliability and validity against the human factors of answering survey questions.

A survey that is too difficult to understand can result in inattentive respondents or skipping questions. Respondents have reported scales containing five, seven, and 10 items as the easiest to use, though they reported greater response accuracy with an even higher number of options (Preston & Colman, 2000). However, scales with more than six options make it more difficult for respondents to differentiate between closely related items, such as the difference between *disagree* and *slightly disagree* (Chang, 1994). Chen et al. (2015) found five anchors to be ideal (compared to four through nine options) as it allowed for quicker interpretation while minimizing bias toward selecting the extreme ends of the scale. Weijters et al. (2010) similarly recommended five anchors, but only when respondents were from the general population. If respondents are known to have high verbal skills and survey experience, they recommend a seven-anchor scale.

Midpoint. The difference between an odd or even number of items in a agree/disagree Likert scale is the inclusion of a midpoint. The impact of a midpoint on reliability and validity is inconclusive (see Krosnick, 1991; see Krosnick & Fabrigar, 1997; see Kulas et al., 2008; see Leung, 2011). Including a midpoint is often preferred, as it allows for a genuinely neutral attitude to be measured instead of forcing respondents to "take sides" (Colman et al., 1997; Johns, 2005). Midpoints also reduce the tendency of respondents to automatically select extreme ends of the scale (Weijters et al., 2010) while minimizing the number of skipped responses (Guy & Norvell, 1977). However, this may indicate noncommittal or satisficing behavior (Matell & Jacoby, 1971). Satisficing behavior occurs when a respondent selects the minimum satisfactory option to complete a task quicker. Researchers concerned with such behavior can mitigate it by including more anchors (Matell & Jacoby, 1971).

Nuances of the midpoint verbiage may also minimize its use as a satisficing option. Choices such as *no opinion* and *unsure* are more likely to garner a midpoint selection than *undecided* and *neither* (Nadler et al., 2015). To balance the desire for high reliability and validity with the equally vital need to make the survey easy to use, the research used a five-anchor midpoint scale (strongly disagree, disagree, undecided, agree, strongly agree).

Ethical Considerations

Anonymity and Confidentiality

The researcher did not collect any personally identifiable information. Such information includes names, phone numbers, physical and email addresses, social security numbers, or Amazon account usernames. However, Amazon likely collected much of this information on users and was aware of which users participated via internal coding. The researcher, however, did not solicit any personally identifiable information from Amazon and assumed participants had accepted Amazon's practice. The researcher relied on MTurk's internal coding to ensure participants in the pilot study were excluded from the main study.

Informed Consent

Before beginning the survey, each participant was provided an electronic informed consent form with the option to agree or disagree. The form provided broad information regarding the purpose of the research, potential discomforts, benefits, confidentiality of records, compensation, researcher contact methods, voluntary participation information, and a consent statement (Embry-Riddle Aeronautical University [ERAU], 2023a). Participants selecting *agree* were directed to the beginning

of the survey. Those participants selecting *disagree* were thanked for their consideration and excluded from the remainder of the survey.

Analysis and Reporting

The researcher incorporated, summarized, and reported all usable data regardless of implication. Results include a summary of demographic data, but no participant-specific data is published. To facilitate the multi-group analysis, the analytic software created two separate groups, but further division based on demographic data did not occur. The researcher reported results without bias or favoritism.

Institutional Review Board

In order to safeguard the rights and safety of participants, the researcher submitted a research application to the Embry-Riddle Aeronautical University IRB for review and approval before recruiting any participants. The IRB uses the *Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research* as its guide (ERAU, 2023a).

The IRB classifies research into three categories—exempt, expedited, and full—which dictates the level of detail in its review. Exempt research encompasses projects presenting minimal risk to participants and involves survey procedures or collecting benign behavioral intentions through written responses (ERAU, 2023b). This study met the criteria of exempt research (see Appendix A).

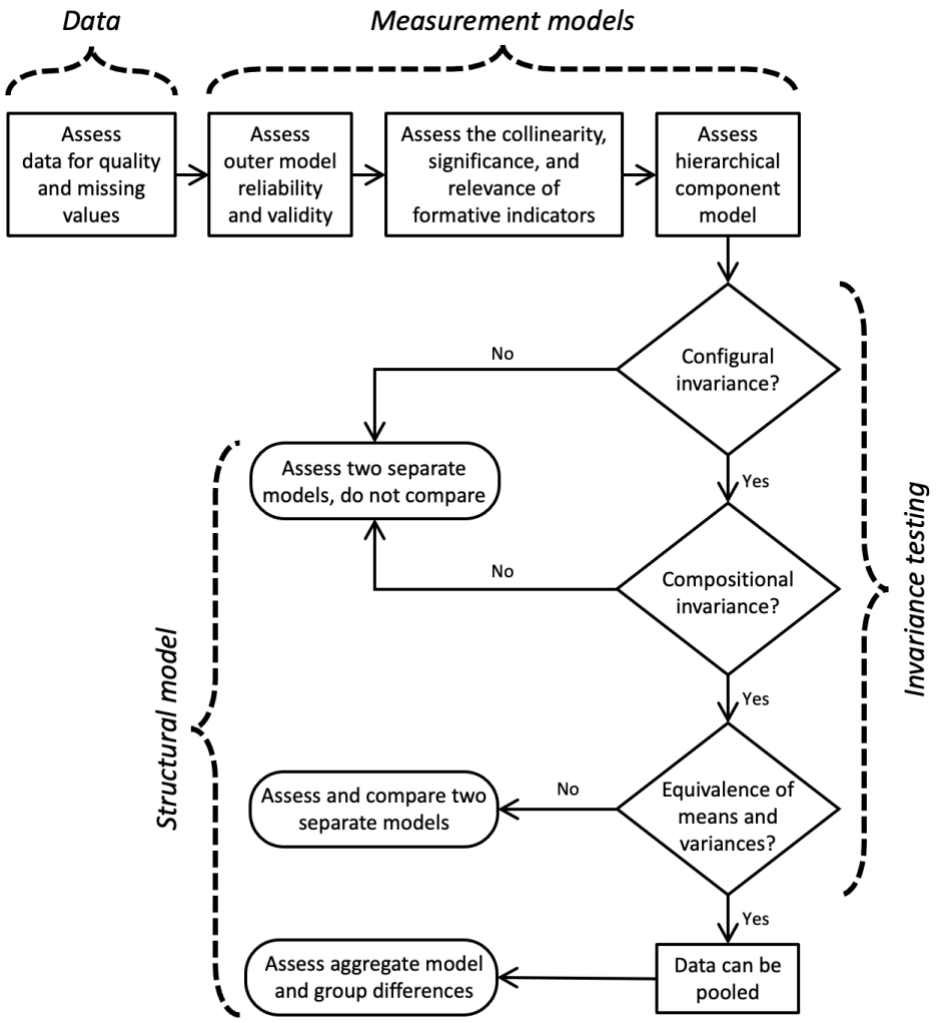
Data Analysis

The major phases of data analysis included assessing data quality, evaluating the measurement models, testing for measurement invariance, and assessing the structural model. Within each of these phases were several unique steps. The final assessment of

the structural model was partially dependent on the results of the invariance testing of two groups (males and females). Figure 4 provides an overview of the data analysis plan, including major phases and steps.

Figure 4

Data Analysis Flowchart Showing Major Phases and Steps



Note. The invariance testing steps constitute the *measurement invariance of composite models* (MICOM) procedure (Hair et al., 2018, 2022; Henseler et al., 2016).

Initial data analysis included compiling descriptive statistics (demographic data), evaluating the usefulness of the data, and managing missing data. Descriptive statistics included age, income, gender, ethnicity, education level, and marital status. Other than the multigroup analysis of gender, the researcher reported, but did not perform, any statistical analysis of demographic data.

Data Quality

Evaluating data quality involved analyzing participants' responses, while managing missing data involved imputing unanswered questions. The researcher reviewed each participant's responses (using Excel and Smart PLS) for suspicious patterns such as straight-lining, diagonal-lining, alternating extremes (Hair et al., 2022), or surveys completed in an inexplicably short amount of time. Such evidence indicates responses with a purpose other than truthfulness (known as *satisficing*). Samples displaying such evidence were removed, but only if the evidence was unambiguous.

Missing Data

The researcher examined missing reflective construct data and managed it by imputation using the inference method (see Hair et al., 2010)—replacing the missing data with the most likely response based on the remaining construct responses (if the remaining responses are highly similar). For example, if a participant omitted one response from the *perceived social influence* questions but answered three with *agree* and one with *strongly agree*, the researcher replaced the missing value in that case with *agree* since it was the most likely response the participant would have given based on the other responses. The method was not utilized for formative constructs since response similarity is not a characteristic.

Formative constructs presented a unique situation regarding missing values. Since indicators are not interchangeable, one cannot accurately apply the inference method based on associated values. In other words, identical responses to any or even all indicators in a formative model are not necessarily evidence of any particular response to a missing value. However, the literature lacks unique methods to handle missing formative data and only references the inference method (see Hair et al., 2010). While mean value imputation is an option, using such a strategy incurs not only distortion of response distribution and depressed correlations (Hair et al., 2010) but would have resulted in differing replacement techniques for different indicators. Imputation using similar participants was another option, but the researcher was concerned about introducing bias due to attributing a particular response based on membership in a demographic group. Because of these disadvantages, the researcher rejected the mean replacement value and demographic-based options and instead adopted a technique traditionally used for ordinal data—the median imputation method (see Alam et al., 2023).

The median imputation method replaces missing values with the middle value of a scale (Alam et al., 2023). The method is simple, but should only be used when the amount of missing data is less than 5% (Alam et al., 2023). In this research, the amount of missing formative data was extremely low (.2%). As a result, this technique was adopted by replacing missing formative values with “3—undecided”. After addressing missing data, the instrument’s reliability assessment followed.

Reliability Assessment of Reflective Constructs

During both the pilot study and main study, the researcher evaluated the reliability of the measurement instrument. Reliability indicates the repeated ability of the instrument to return identical values over multiple iterations. It is often confused with accuracy but differs in that accuracy is an instrument's ability to return the true value, while reliability is an instrument's ability to return the same value. Although both are equally important, the researcher only evaluated reliability. Accuracy was assumed since it relates to the true attitudes and characteristics of the participant. Evaluating the reliability of reflective constructs involved assessing indicator reliability and internal consistency reliability, of which there are three techniques.

Indicator Reliability. To evaluate indicator reliability of reflective constructs, the researcher examined the outer loadings of each indicator. Each of these loadings should be statistically significant ($p < .05$). Also, standardized outer loadings (l) should be at least .708 since that value indicates that at least 50% of the variance of an indicator is explained by the construct (l^2 = % of variance explained by construct) (Hair et al., 2022). An outer loading less than .708 indicates more of an indicator's variance is explained by its error than by the construct. The researcher considered removing indicators with a loading less than .708, but only if doing so increased the internal consistency reliability or convergent validity (Hair et al., 2022). Indicators with outer loadings less than .4 were considered for removal outright (see Hair et al., 2022).

Internal Consistency. Internal consistency indicates how well a construct's measures correlate with each other (Hair et al., 2022). Since the measures of a reflective construct measure the same phenomenon, they should have similar values. Three options

to measure internal consistency are Cronbach's alpha, composite reliability, and the reliability coefficient. To provide full disclosure of reliability, the researcher computed and reported results from all three methods.

Cronbach's Alpha. Cronbach's alpha is based on the intercorrelations of a construct's indicators. The generally accepted lower limit is α = .7, though α = .6 (see Van Griethuijsen et al., 2014) or even α = .5 (see Hair et al., 2010) is also accepted by some researchers. Values above α = .9 generally indicate redundant items and should be considered for removal (Hair et al., 2022). All reflective constructs were sourced from literature, and Cronbach's alpha from their associated research is shown in Table 3.

Table 3

Reported Reliability of Source Measures (Reflective Constructs Only)

Construct	α	Source
Recognizing the need for intervention	.88 - .93[a]	Albayrak-Aydemir and Gleibs (2021)
Taking responsibility	.79	Nickerson et al. (2014)
Knowing what to do	.71 - .77[a]	Albayrak-Aydemir and Gleibs (2021)
Deciding to act	.70	Burn (2009)
Out-group social bias	.85[b]	Abbott and Cameron (2014)
Situation outcome	.75	Pöyhönen et al. (2012)
Victim outcome	.78	Pöyhönen et al. (2012)
Self-efficacy	.90[c]	Schwarzer (1999)
Perceived social influence	.69	Pozzoli and Gini (2010)

[a] Albayrak-Aydemir and Gleibs (2021) included reliability scores for three independent samples. The values for *recognizing the need for intervention* were $\alpha = .88, .89.$ and .93. For *knowing what to do*, the values were $\alpha = .71, .74.$ and .77.

[b] Abbott and Cameron (2014) did not explain reporting a Cronbach's alpha despite their *social bias* construct appearing to be formative.

[c] Schwarzer (1999) did not include an American sample, so the British sample's reliability is provided to maintain language continuity.

Although it is arguably the most widely used, Cronbach's alpha is sensitive to the number of observed variables (Hair et al., 2022), favoring constructs with fewer. It also assumes all indicators are equally reliable (via loadings) (Hair et al., 2022) and equally covary (Dijkstra & Henseler, 2015). Neither of these assumptions may be true, nor do they align with the PLS algorithm, since PLS assumes neither equal loading nor equal covariance of indicators. These assumptions generally result in underestimating an instrument's internal consistency reliability (Hair et al., 2010, 2022). In order to capitalize on PLS-SEM's prioritization of indicators relative to their loading, many researchers prefer composite reliability over Cronbach's alpha.

Composite Reliability. Composite reliability does not assume equal outer loading, making it more applicable to PLS-SEM methods. It derives its value via the individual loadings and indicator error (see Hair et al., 2022), with higher loadings and smaller error returning a greater reliability quotient. However, the formula uses standardized loadings, which may be inconsistent with some research designs (such as this one), which use unstandardized coefficients. Acceptable values for composite reliability are the same as

those of Cronbach's alpha. However, whereas Cronbach's alpha may underestimate reliability, composite reliability tends to overestimate it (Hair et al., 2022). Some researchers have begun using the reliability coefficient as a compromise.

Reliability Coefficient. In an effort to develop a reliability test better suited to PLS, Dijkstra and Henseler (2015) developed the reliability coefficient. The method relies on indicator weights (the strength of the effect of the construct on the indicator) instead of loadings (the level of correlation between a construct and indicator) (Dijkstra & Henseler, 2015). The reliability coefficient tends to fall between Cronbach's alpha and composite reliability, and many PLS users consider it a superior reliability measure (Hair et al., 2022). To report reliability as fully as possible, Cronbach's alpha, composite reliability, and the reliability coefficient were all reported.

Reliability Assessment of Formative Constructs

Traditional methods to evaluate reliability (such as those above) are inappropriate for formative constructs (Hair et al., 2018). That fact is due to the lack of a requirement for collinearity among formative indicators. Test-retest is the only effective way to truly measure the reliability of a formative construct (Hair et al., 2018). However, due to the threats to validity, the desire to preserve anonymity, and the general difficulty of retesting participants, an evaluation of the reliability of formative constructs was omitted (see Hair et al., 2018).

Validity Assessment of Reflective Constructs

Convergent Validity. Since all indicators in a reflective measurement model are expected to converge on the same latent variable, they should all be highly correlated. The level of these correlations indicates the degree of convergent validity. Commonly

used statistics are the Average Variance Extracted (AVE) and factor loadings. Factor loadings are specific to each measure, while AVE represents an average of all the measures in the outer model. Factor loadings of $l > .5$ indicate acceptable loading (> 50% correlated variance), though $l > .7$ is generally preferred (Hair et al., 2010). In addition, the factor loadings of an indicator and its construct should be greater than their loading on any other construct. While factor loadings measure each indicator in isolation, AVE examines all of a factor's loadings in aggregate.

AVE describes how much indicator variance is explained by the construct and how much is explained by other influences (manifested as error). It is not indicator-specific but rather the average variance explained for all the indicators of a construct. A value of .3 means that, on average, the construct explains 30% of the variance of the indicators. Generally, an AVE of .5 or greater is preferred (Hair et al., 2010). A value less than .5 would indicate that more variance is due to error than the construct.

Discriminant Validity. Discriminant validity is the degree to which a construct's operationalization is dissimilar to other constructs with which it should theoretically not be similar (Hair et al., 2022). In other words, a construct should be unique, both in terms of how it relates to other constructs and how its indicators relate to it. To test for discriminant validity of reflective constructs (including the lower-order constructs *situation outcome* and *victim outcome*), Smart PLS software compares the average variance extracted (AVE) of a construct with the square of the correlation between it and each of the other constructs (the Fornell-Larcker criteria; Hair et al., 2022).

An AVE greater than the square of the correlation between constructs (the percent in common) indicates that the construct shares more variation with its indicators than it

does with any other construct in the model (Hair et al., 2010). Conversely, if a construct shares more variance with an unrelated construct (square of the correlation) than it does with its indicators (AVE), it is more similar to the other construct than unique. However, the Fornell-Larcker criteria has drawn criticism due to its inability to accurately describe the discriminant validity of constructs whose indicators have similar loadings (within .2 of each other) (Henseler et al., 2015). An alternate test of discriminant validity is the heterotrait-monotrait ratio (HTMT).

The HTMT compares the average correlation of indicators to their unpartnered constructs against the average correlation with which the indicators are assigned (see Henseler et al., 2015). The method is more easily relatable (particularly during CFA), as it uses correlation values (the essence of discriminant validity) instead of loadings or AVE, which are arguably indirect measures. However, there is no agreed-upon threshold for HTMT. Some researchers propose a value of .9 as an upper limit (see Clark & Watson, 1995) for models containing similar constructs (Hair et al., 2022), while others propose a more rigorous .85 as an upper limit (see Kline, 2011) for models whose constructs are more obviously distinct (Hair et al., 2022). To report discriminant validity as fully as possible, the researcher reported the results of both the above methods.

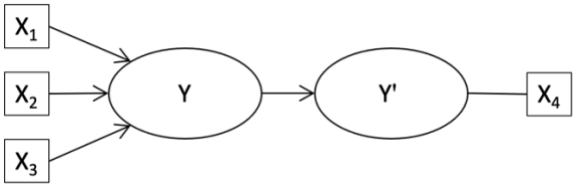
Validity Assessment of Formative Constructs

Convergent Validity. Since measures of a formative construct are not expected to correlate, researchers evaluating convergent validity perform redundancy analysis (Chin, 1998), which evaluates how well a formative construct correlates with the same construct measured reflectively. One method to do this is to construct a global item that captures the essence of the construct measured by the formative indicators (Sarstedt et al.,

2019). The global indicator is assigned to a reflective construct (established for the sole purpose of validity assessment) whose correlation with the original construct is then measured. Convergent validity can then be analyzed by establishing the level of correlation between the two constructs. Since conceptually, the two constructs represent the same thing, they should positively correlate with each other. Figure 5 illustrates an example temporary model created to measure convergent validity.

Figure 5

Example of a Formative Construct (Y) and Single Global Item Reflective Construct (Y')



In Figure 5, Y is a construct measured formatively, while Y' is the same construct measured reflectively. Since Y and Y' are the same construct, differing only in their measurement theory, they should be highly correlated. Although single-item constructs are generally considered insufficient for model estimation, their use as a tool for evaluating convergent validity is accepted since they are only temporary surrogates used for this purpose (Hair et al., 2022). Since the two constructs are operationally identical (even though their measurement theories differ), their relationship should indicate high covariance (path coefficient) and explained variance (R^2 of Y'). A path coefficient of $\beta >$ .7 and an $R^2 >$ .64 (or at least $R^2 >$ .5) indicate convergent validity (Hair et al., 2022). The researcher applied this method to assess the convergent validity of the constructs

nonintervention cost, *out-group social bias*, *intervention skills*, and *expectation of positive outcome*.

Content Validity. Content validity is the extent to which observable measures are relevant to and an accurate representation of their target latent construct (Almanasreh et al., 2019). The only outer model absent in the literature was *nonintervention cost*, necessitating an evaluation of the content (measures) assigned to represent it. Unlike other types of validity, the evaluation of content validity is primarily subjective. Researchers commonly rely on subject-matter experts (SMEs) to assess the relevance and accuracy of the indicators to the construct. However, the difficulty lies in identifying the level of agreement required to establish an adequate level of content validity. Statisticians have thus attempted to quantify SME subjectivity.

Individual Content Validity Index. A popular approach to quantifying content validity is to compute each indicator's item content validity index (I-CVI). Its value is the proportion of SMEs who assign the relevance of the indicator in the positive half of a Likert scale measuring the item's importance:

$$I\text{-}CVI = A \div N \tag{3}$$

Where:

A = the number of SMEs who scored the indicator as relevant

N = the total number of SMEs

Although a scale of four anchors is most commonly used (Almanasreh et al., 2019), researchers can use any number of anchors since they ultimately convert them into the binary choice of *relevant* or *not relevant*. In other words, SME ratings are converted into whether they selected either the upper or lower half of the Likert scale. Although offering

a scale greater than two anchors adds no statistical value, it may enable SMEs to formulate their responses better.

The minimum I-CVI needed to justify retaining an indicator is *I-CVI* = .78 (Polit et al., 2007). However, that standard heavily biases I-CVI toward exclusion. For example, with a group of nine SMEs, only two votes for *not relevant* override the *relevant* votes of the other seven. Also, reducing the Likert scale to a binary response makes its value susceptible to being amplified purely based on chance (Wynd et al., 2003), particularly with a small number of SMEs. As a response, Polit et al. (2007) adjusted I-CVI to remove the statistical increase due to chance responses and provided categories of acceptability.

Modified Kappa Index. The modified kappa index (k^*) is similar to I-CVI in its attempt to quantify a subjective assessment by SMEs. However, it refines I-CVI by adjusting its value to account for the possibility that the .78 threshold was reached purely by a chance response (Almanasreh et al., 2019). Of course, the possibility of this occurring should be slight, as one would expect SMEs to provide an informed assessment instead of a random one (Polit et al., 2007). Nonetheless, the adjustment adds a level of rigor which in and of itself increases evidence of validity. The value of k^* is (Almanasreh et al., 2019):

$$k^* = (I\text{-}CVI - p_c) \div (1 - p_c) \quad (4)$$

Where

$$p_c = [N! \div A!(N - A)!] \, .5^N \quad (5)$$

And

A = the number of SMEs who scored the indicator as relevant

N = the total number of SMEs

As with I-CVI, the Likert scale is collapsed to the binary *relevant* or *not relevant*, so researchers can select any number of Likert anchors to better facilitate accurate responses. As the number of SMEs increases, the possibility of agreement due to chance decreases, resulting in k^* and I-CVI converging when the group size reaches approximately 10 SMEs (Polit et al., 2007). Polit et al. (2007) combined several standards (see Cicchttie & Sparrow, 1981; see Fleiss, 1981; see Landis & Koch, 1977) and assessed various combinations of N and A to propose three levels of content validity: *fair* (.40 - .59), *good* (.60 - .74), and *excellent* (> .74). The value of using such levels is that it allows recognition that validity itself is not a binary *valid* or *not valid* concept, but instead one with varying degrees.

Selection of Subject Matter Experts. SMEs should understand both the construct meaning and measurement theory enough to allow them to make an informed assessment while understanding the purpose of doing so. The selection of SMEs is driven less by concern for generalizability or representativeness and more by expertise, quality of input, and trusted dedication to the task. The size of the SME group is usually relatively small—between five and 10 experts (Almanasreh et al., 2019).

Understanding the construct *nonintervention cost* necessitates experience with commercial air travel from the passenger perspective, while understanding measurement theory comes from academic training in SEM. To assemble SMEs with both of these requirements, the researcher elicited support from eight aviation experts who are experienced with commercial air travel and have successfully completed academic SEM coursework. The commercial air travel experience allowed them to conceptualize the

potential individual cost due to an attack on a flight attendant. Completing academic coursework in SEM allowed for understanding content validity and measurement theory.

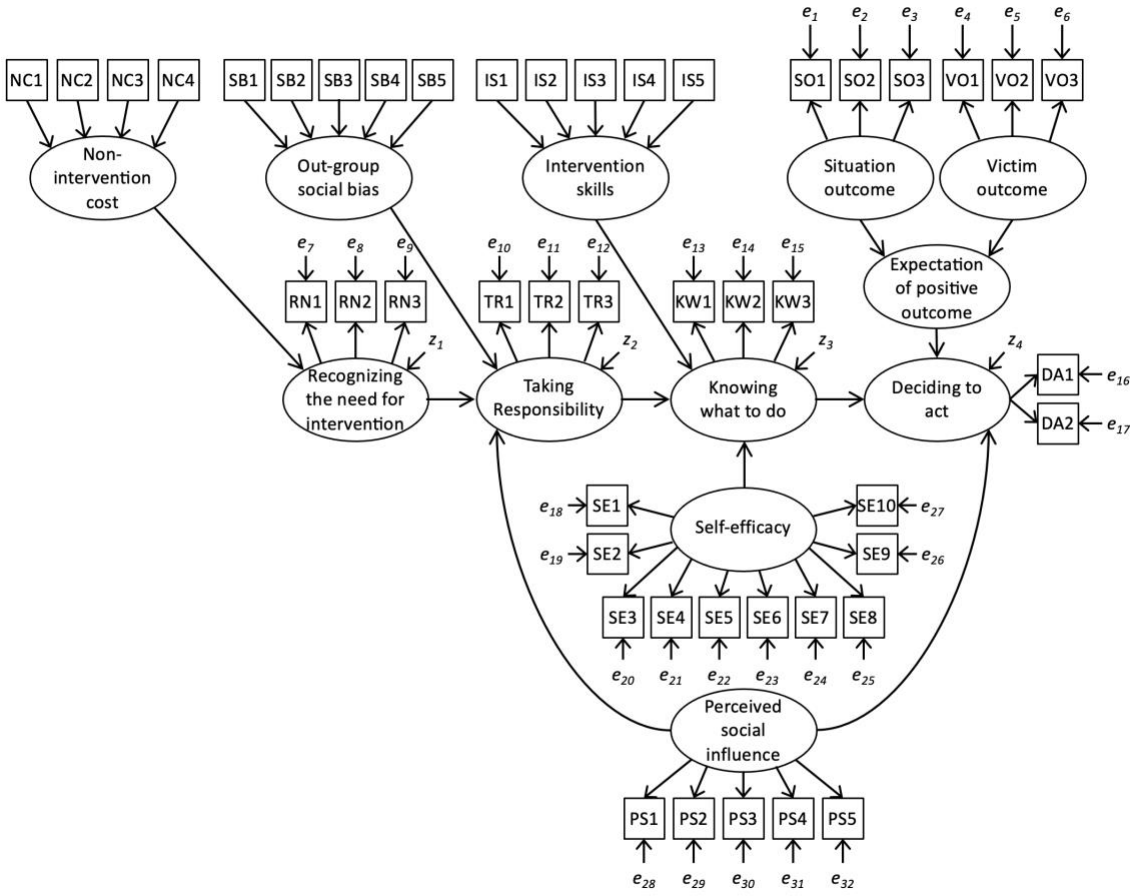
Discriminant Validity. Discriminant validity is the degree to which a construct's operationalization differs from other constructs. In other words, a construct should be unique, both in terms of how it relates to other constructs and how its indicators relate to it. Since traditional methods to evaluate discriminant validity rely on fit indices, they are incompatible with formative constructs in PLS-SEM, which does not use fit indices during model estimation. Researchers have yet to develop accepted methods to test for discriminant validity of formative constructs (Wang et al., 2015). Thus, the discriminant validity of the three formative constructs was left unevaluated.

Assessment of the Structural Equation Model

Evaluating the full model involved assessing outer models followed by the inner model, with the single hierarchical model also necessitating unique analysis. Figure 6 shows the entire model, including indicators and errors.

Figure 6

The Full Research Model



Note. The global indicators (NC5, SB6, IS6, and EO1) and repeated indicators of *expectation of positive outcome* are not shown. In PLS-SEM, formative constructs are considered to be error-free composites ($z = 0$) and reflective error variables, shown for understanding only, are not specified.

Measurement Models. In addition to reliability and validity analysis (accomplished earlier), the indicators of formative constructs were assessed for collinearity and relevance. There were no additional assessments (other than reliability and validity) of reflective measurement models.

Collinearity of Formative Indicators. Since formative indicators measure different dimensions of their construct, they should not covary. Doing so indicates they more closely measure the same dimension instead of different ones and manifests as misleading indicator weights during model estimation (Hair et al., 2022). The variance inflation factor (VIF) was used to assess collinearity.

To compute VIF, each indicator (*i*) is regressed onto the remaining indicators of the construct to establish the level of explained variance (Hair et al., 2022). The formula for VIF is:

$$VIF_i = \frac{1}{1 - R_i^2} \quad (6)$$

Where

R_i^2 = the variance of indicator *i* explained by its co-indicators

Ideally, VIF is less than *VIF* = 3 (Hair et al., 2022), which equates to 67% of an indicator's variance being explained by its co-indicators. Indicators with *VIF* > 5 became candidates for removal (Hair et al., 2022).

Significance and Relevance of Formative Indicators. Formative constructs in PLS-SEM are assumed to be composites wholly bounded by their indicators (the explained variance of an exogenous formative measurement model is always 1 (Hair et al., 2022). In some cases, however, an indicator will contribute little to forming the construct. In such cases, researchers should consider removing the indicator from the measurement model (Hair et al., 2022). The weights and loadings of formative indicators were examined to identify any that were not statistically different from zero.

The outer weight of a formative indicator represents its relative contribution to forming the construct (Hair et al., 2022). Since outer weights are relative values, the

maximum and average outer weight depends on the number of formative indicators (Hair et al., 2022). Therefore, there is no minimum value required to justify retainment. Instead, statistical significance becomes the standard. Indicators with nonsignificant outer weights were considered for removal, but only if their absolute contribution (loading) was also small (Hair et al., 2022).

A formative indicator's loading represents the absolute (nonrelative) contribution to forming the construct (Hair et al., 2022). Even if an indicator's relative contribution is not significant, its absolute contribution (measured independently from its co-indicators) may be enough to warrant retention. That possibility is especially likely for constructs with many indicators (Hair et al., 2022). Indicators with an outer loading less than $l = .5$ were considered for removal, especially if the outer loading was also not statistically significant (Hair et al., 2022).

Hierarchical Component Model. The construct *expectation of positive outcome* is a hierarchical component model (HCM). The two lower-order constructs (LOCs), *situation outcome* and *victim outcome*, each comprise reflective measurement models. In contrast, the higher-order construct (HOC) *expectation of positive outcome* comprises a formative measurement model. In its entirety, the trio comprises a reflective-formative HCM. As such, it required a somewhat unique assessment.

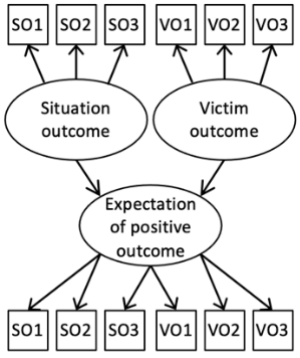
The reliability and validity of the two LOCs were assessed the same as other reflective constructs (see Hair et al., 2018). However, the two LOCs also exist as formative indicators of the HOC. That fact necessitated evaluating the collinearity of the LOCs (via VIF) as well as the significance and relevance of their relationship to the HOC (using path coefficients as outer weights) in the same manner as other formative

constructs (see Hair et al., 2018). Also, since PLS algorithms require at least one indicator be assigned to each latent construct, the repeated indicators method was used (see Hair et al., 2018).

Compared to other methods, the repeated indicators method produces better estimations of the intra-HCM paths (from the LOCs to the HOC) (Sarstedt et al., 2019). The method entails assigning all LOC indicators to the HOC using the same measurement theory used in the LOC measurement models. In other words, since the LOC indicators are reflective, the repeated indicators were also reflective. However, HOC model assessment is only relative to its LOCs and not the repeated indicators. Special consideration is also required if the LOCs have different numbers of indicators or if the HOC has predictor constructs (Sarstedt et al., 2019), though neither condition was present in the research model. Figure 7 illustrates the *expectation of positive outcome* HCM with repeated indicators.

Figure 7

Hierarchical Component Measurement Model With Repeated Indicators



Multi-group Analysis. Included in the sample were two distinct groups—males and females. The ambiguous findings regarding gender's influence on intervention means heterogeneity was highly possible. That possibility warranted special consideration for the two groups and had two possible outcomes. If, during multigroup analysis (MGA), no differences were discovered (homogeneous groups), the data for the two groups would have been combined, and a single structural model assessed. However, since heterogeneity was confirmed, invariance testing determined the treatment of the structural model.

MGA determines if the path coefficients for the two groups are statistically different from each other. For analysis of only two groups, two types of MGA are available—parametric and nonparametric, although researchers have found the former to be prone to type I errors (Hair et al., 2018; Sarstedt et al., 2011). Within nonparametric tests, there are two options—PLS-MGA and permutation.

PLS-MGA. PLS-MGA relies on samples of target groups to evaluate the probability (*p*) that in the population, the path coefficient found in the first group is larger than that found in the second group ($\beta^1 > \beta^2$). The method does so by creating a bootstrap sample from the members of one target group, estimating the path coefficient, and comparing it to the path coefficients of multiple (> 5,000) bootstrap samples from members of the second target group (see Henseler et al., 2009). The objective is to determine the probability that $\beta^1 > \beta^2$. However, the method is one-sided, as it cannot perform a two-tailed test (Hair et al., 2018; Henseler et al., 2009). Researchers who wish to evaluate the other direction ($\beta^2 > \beta^1$) must manually *inverse* the probability statistic and rerun the analysis (see Hair et al., 2018). Doing so, however, introduces the

possibility of inadvertently corrupting the data and generating inaccurate results. Researchers who wish to conduct a full two-tailed analysis can also use an alternate approach, such as permutation.

Permutation. The permutation method evaluates if the path difference between two target groups is statistically different from the path difference between two other randomly-generated groups from the full dataset. The method involves three steps (see Hair et al., 2018; see Sarstedt et al., 2011):

1. Estimate the models for each target group (M = males, F = females) to obtain the path coefficients (β) and compute the difference between them:

   $$d^{MF} = |\beta^{M} - \beta^{F}| \quad (7)$$

2. Randomly permutate the full dataset over numerous iterations ($i > 1{,}000$) to create new pairs of groups (A_i and B_i) and obtain the difference between the path coefficients:

   $$d_i^{AB} = |\beta_i^{A} - \beta_i^{B}| \quad (8)$$

3. Create a two-tailed 95% confidence interval by sorting the resulting permutated path differences and obtain boundaries of the middle 95% (between the 2.5% and 97.5% values). If the difference in path coefficient of the targeted groups falls within the permutated boundaries, its difference is not significant, and the groups are homogeneous:

   $$d_{2.5\%}^{AB} < d^{MF} > d_{97.5\%}^{AB} \quad (9)$$

In other words, the difference between the two original groups is no different from that between 95% of the randomly-generated groups. However, a significant difference outside these boundaries indicates that the groups are heterogeneous.

A limitation of the permutation approach is that it necessitates equal group sizes (Hair et al., 2018; Sarstedt et al., 2011). Such a requirement may be problematic when conducting MGA on datasets collected without stratification. In such an event, researchers should randomly eliminate observations from the larger group to obtain equal sizes or select another MGA method (Hair et al., 2018). The present research used the permutation method for MGA and stratified data collection to obtain equal group sizes (males and females). Since sizes became unbalanced due to data quality issues, the researcher randomly eliminated observations from the larger group.

Invariance Testing. Before performing a meaningful comparison of the two groups, the researcher first established that any variance between them was due to true heterogeneity, not how they were measured or their interpretation of constructs. Testing this invariance was via a procedure developed for composite models (such as PLS)—the *measurement invariance of composite models* (MICOM) procedure. The procedure involves establishing three types of variance: configural, compositional, and equality of composite means and variances (Henseler et al., 2016). Achieving all three types establishes full measurement invariance, enabling group-specific structural model relationships (path coefficients) to be meaningfully compared (Hair et al., 2018, 2022; Henseler et al., 2016). It also enables the data from the groups to be combined or pooled, resulting in added statistical power and additional options for follow-on analysis (Hair et al., 2018, 2022; Henseler et al., 2016). Achieving only the first two types (configural and compositional) establishes partial measurement invariance and allows for group comparison but not data pooling (Hair et al., 2018, 2022; Henseler et al., 2016). Only achieving or failing to achieve the first type (configural invariance) means the models of

the two groups must be considered independently without comparison or pooling (Hair et al., 2018, 2022; Henseler et al., 2016).

Configural Invariance. The first step in the MICOM procedure is establishing that the measurement models are specified equally across groups (Hair et al., 2018, 2022; Henseler et al., 2016). The step is the only one that is qualitative, as it does not rely on any statistical analysis. Instead, it requires a simple verification of identical indicators and data treatment (Henseler et al., 2016). The latter includes the strategies for treating missing data and any algorithm settings used during estimation. Establishing configural invariance can be threatened when combining data with that from prior research, though such is not the case with this research.

Compositional Invariance. The second step of the MICOM procedure, establishing compositional invariance, involves a quantitative assessment centered on the outer weights. It establishes that the composite (construct) scores have been created the same across groups (Henseler et al., 2016). Establishing configural invariance can be problematic when using cross-cultural or multi-lingual data, though neither is present in this case.

One option to ensure compositional invariance is to utilize fixed indicator weights during specification. Doing so will ensure compositional invariance since it will have been specified in the design (Henseler et al., 2016). However, the option requires assigning one indicator in each outer model as the dominant indicator—having the largest dimensional contribution (Henseler et al., 2016). Using this option incurs risk, however, as the researcher must establish the lead indicator through literature. Doing so can lead to misidentifying the true dominant indicator, resulting in inaccurate results.

Henseler et al. (2016) provide an alternate permutation test if fixing indicator weights is not desired. Their process involves first identifying the outer weights (ω^M and ω^F) and composite scores (ξ) of the two groups (*M* and *F*). If the two groups are invariant, their correlation will have a value of one (or *not* statistically different from one). Using the outer weights obtained from *M* and *F*, the researcher applies them to two randomly created groups (*A* and *B* from the entire dataset) to obtain new construct scores and their level of correlation (-1 to 1). The process repeats multiple times (> 5,000).

$$c = \text{cor}\ (\xi^M \omega^M\ \xi^F \omega^F) = \text{cor}_i\ (\xi^A \omega^M\ \xi^B \omega^F)_i \quad (10)$$

Where

c = correlation between composite scores

i = iteration

Obtaining construct scores (and their correlation) for each iteration creates a distribution of correlation values. If 95% of the distribution of *c* is not statistically different from a value of one ($p \prec .05$), compositional invariance is established (Henseler et al., 2016). Achieving the second level of invariance (partial invariance) allows the estimation results of the two groups (R^2 and path coefficients) to be meaningfully compared to each other (Hair et al., 2018, 2022; Henseler et al., 2016).

Equality of Composite Means and Variances. The final step of the MICOM procedure is evaluating the equality of mean scores and variances for the two groups. Establishing this equality results in full measurement invariance and the pooling of data for single-model assessment (Henseler et al., 2016). The method is straightforward and similar to the previous step. The mean construct scores and variances from one group (males) are compared to the mean construct scores and variances from the second group

(females). Additional multiple permutations (> 5,000) are conducted on randomly-generated groups. Full measurement invariance is established if the two means and variances are not statistically different (at the 95% confidence level). The data can then be combined into a single model for assessment. Researchers, however, should take caution in reporting the results of such a model.

Single-model assessment affords greater statistical power but does not discount that heterogeneity exists. On the contrary, full measurement invariance merely allows a single model to be created for later moderation analysis of any heterogenous groups. If MGA identifies structural differences (path coefficients) between groups, then pooling the groups into one aggregate model without some allowance for heterogeneity will lead to inaccurate conclusions (Henseler et al., 2016). This research only established whether or not males and females were heterogeneous in the assault context and model and reported differences between the two groups.

Structural Model. Assessing the structural model involved examining four characteristics: collinearity, significance and relevance of relationships, explanatory power, and predictive power (Hair et al., 2022).

Collinearity of the Structural Model. In much the same way formative indicators will be assessed for collinearity, so too will co-predictor constructs. For instance, in the structural model (see Figure 3), the constructs *out-group social bias*, *recognizing the need for intervention*, and *perceived social influence* are co-predictors of *taking responsibility*. High levels of collinearity between any pairs can result in incorrect path coefficients or even erroneous directional relationships (incorrect assignment of a positive or negative sign to the path coefficient) (Hair et al., 2022). A collinearity evaluation of all of the

following pairs was conducted (see Table C1 for illustrative partitions): (a) *out-group social bias* and *recognizing the need for intervention*, (b) *out-group social bias* and *perceived social influence*, (c) *recognizing the need for intervention* and *perceived social influence*, (d) *intervention skills* and *taking responsibility*, (e) *intervention skills* and *self-efficacy*, (f) *taking responsibility* and *self-efficacy*, (g) *expectation of positive outcome* and *knowing what to do*, (h) *expectation of positive outcome* and *perceived social influence*, (i) and *knowing what to do* and *perceived social influence*. If constructs displayed excessive collinearity, the researcher considered three options: eliminating a construct, combining indicators, or establishing HOCs (see Hair et al., 2022).

Significance and Relevance of Relationships. There are two methods to evaluate the significance of path coefficients: *t* and *p* values or confidence intervals (Hair et al., 2022). The first method involves evaluating the construct relationships (path coefficients) for statistical significance by examining *t* values (as in *t*-test) at the selected significance level (*p*). A *t* value greater than the two-tailed critical value (1.96 for $p = .05$) indicates the relationship is statistically different from zero (i.e., zero means no relationship; Hair et al., 2022). Using a two-tailed test supposes that the direction (sign) of the relationship could be either direct (positive) or inverse (negative). The positive or negative direction is not to be confused with the direction of the path between constructs. Instead, it relates to the direction of variance (increase or decrease) of the endogenous variable relative to the direction of variance of its predictor. While the method is acceptable, it does not indicate the stability of the path coefficient or how close the coefficient plus error may be to zero (Hair et al., 2022). Examining the bootstrap confidence interval adds greater clarity to that unknown.

The bootstrap confidence interval is the range a researcher can claim (with the selected confidence level) that the true path coefficient will fall (Hair et al., 2022). In other words, a bootstrap confidence interval of *A* to *B* using $p = .05$ means a researcher can claim with 95% certainty that the true path coefficient falls somewhere between *A* and *B*. If $A \neq 0$ and $B \neq 0$, the true path coefficient (falling between *A* and *B*) is significantly different from zero. The added benefit of using this method lies in the range of bootstrap differences between *A* and *B* (Hair et al., 2022). A wide range indicates less stability and lower accuracy in out-of-sample (the population) path coefficient predictions. A narrow range indicates more ability to predict the true path coefficient in the population accurately. Due to the greater clarity it provides, this method was used to evaluate the significance of the model's path coefficients.

Explanatory Power. The explanatory power of a model represents the ability of the exogenous variables to effectively predict the endogenous ones or to explain the model relationships as specified (Hair et al., 2022). The focus of explanatory power is only on the sample data (as opposed to out-of-sample data). In other words, within the confines of the sample, exploratory power answers the question of how well the antecedents predict their target variables. To evaluate explanatory power, the researcher used the coefficient of determination (R^2), or the amount of endogenous variance explained by all its predictor constructs (Hair et al., 2022). The R^2 range is between zero and one, with higher values indicating more variance explained. However, there is no minimum threshold for R^2. Even a small R^2 may be entirely acceptable and valuable in fields such as finance (Hair et al., 2022) or safety. In general, there is no upper limit as well, as simply having a greater number of predictors will inherently predict more

variance (Hair et al., 2022). Having an $R^2 > .9$, however, may indicate that the model is too specific to the sample and, thus, might not be representative of the population. Such models may exhibit poor out-of-sample predictive ability (Hair et al., 2022).

Predictive Power. A model representing the true relationships found in the population should have some ability to predict the observed variables of endogenous constructs. In other words, given the values of predictor indicators, a truly representative model should allow practitioners to predict the target indicators. To assess this, researchers will often separate the dataset into a training sample for estimation and a holdout sample for a one-time assessment of predictive ability. A similar but alternate method is $PLS_{predict}$.

$PLS_{predict}$ separates the dataset into several equal parts and conducts separate predictive runs using a different part as the holdout sample and the remaining parts as the training sample (Shmueli et al., 2019). Often, the process is repeated several times. The method creates a more accurate assessment of predictive power because, throughout the iterations, all parts of the data, at some point, will be used as training and holdout data. A researcher decides how many parts to divide the data into (known as *folds*) and how many times to repeat the entire process. When deciding on the former, researchers need to ensure that the training sample contains enough observations, as dictated by minimum sample size requirements (Shmueli et al., 2019).

In this research, the minimum sample size was $n = 550$ per group. Although more than the minimum was collected, much was unusable due to data quality. The less extra data collected, the more folds required, with 10 typically being the maximum number of

folds and repetitions (Shmueli et al., 2019). In order to select the number of folds that would ensure n = 550 in the training sample, the researcher used the following formula:

$$F = \frac{S}{S-550} \quad (11)$$

Where

F = the minimum number of folds

S = the group (male or female) usable sample size collected

After performing the predicative runs, $PLS_{predict}$ provided, as prediction statistics, the mean absolute error (MAE) and the root mean square error (RMSE). MAE represents the difference between the true observations (the holdout) and what the training data predicted (i.e., small MAE indicates high predictive power; Hair et al., 2022). While it is easy to interpret, MAE may underestimate the impact of large errors since it assumes error equality (Hair et al., 2022). RMSE, on the other hand, represents the square root of the average differences (Hair et al., 2022). Like MAE, small RMSE values indicate high predictive power. However, RMSE differs from MAE because RMSE does not assume error equality. As a result, large errors are weighted appropriately. Such weighting makes RMSE the preferred statistic, though it is much less intuitive and more difficult to understand (Hair et al., 2022).

In order to better understand RMSE (and MAE), Shmueli et al. (2019) suggest comparing both to a simple linear regression model benchmark (LM). LM represents a single endogenous indicator's regression onto all the predictor indicators without consideration of the inner model structure. For example (see Figure 6 for reference), DA1 is regressed onto SO1-SO3, VO1-VO3, KW1-KW3, and PI1-PI5. Since LM does not

consider the underlying structural relationships, it should underperform compared to MAE and RMSE (Hair et al., 2022). Interpreting the results involves comparing values.

Regardless of which one a researcher selects, if the prediction statistic (MAE or RMSE) is smaller than all target indicators' LMs, the model has high predictive power (Hair et al., 2022). Accordingly, the model has medium predictive power if most values are smaller than LM. If only a few values are smaller, it indicates low predictive power, whereas LM exceeding all predictive statistic values indicates a lack of predictive power (Hair et al., 2022).

Summary

Chapter III explained how the research was categorized in terms of method, design, and approach. It explained, in broad terms, the difference between CFA and EFA, as well as CB-SEM and the selected approach, PLS-SEM. The population and sample size section explained how the minimum sample size was computed, the benefits and deficiencies of the many methods, and why the selected method was chosen. An explanation of the data and how it was collected within ethical guidelines followed, including a description of how it was examined and assessed.

The chapter explained how the measurement models were assessed for reliability and validity. Since heterogeneity was a possibility, the method to identify it and determine the level of measurement invariance via the MICOM procedure was explained. The chapter closed with an explanation of the plan to assess the structural model and how heterogeneous groups were compared.

Chapter IV: Results

Chapter IV explains the process used to conduct both pilot and main studies as well as reviews the results. There is a detailed explanation of the data treatment and outer-model analysis during the three iterations and reasons for modifying the structural model. The chapter also explains the assessment of the inner model characteristics, invariance testing, and multigroup analysis. It closes by reviewing the level of support for the hypotheses.

Pilot Study I

Pilot study I was conducted in August 2023 to assess the operability of the survey instrument with a secondary objective of identifying opportunities to improve reliability and validity. The initial dataset consisted of 103 unstratified participants. Even though the researcher requested 100 workers via the MTurk interface, 103 MTurk workers submitted surveys. The discrepancy (3% overage) is likely due to workers completing the survey but failing to submit the work in MTurk. All participants who completed the MTurk work submission were compensated per IRB policy.

The researcher used the Microsoft Excel® STDEV, and COUNTIF features to identify straight-lining and inattentiveness. The features return the standard deviation and quantity of particular responses for each participant. A standard deviation of zero occurs if all answers are the same. Counting the number of each response reveals occasions where only one answer is different.

Three participants were removed due to straight-lining (all responses the same) and two more for near straight-lining (answering 56 of the 57 responses the same). Eleven more participants were removed due to having a high probability of being non-

human (Qualtrics CAPTCHA score < .5). Table 4 displays the demographics of the remaining sample (n = 87). Since the purpose of the pilot study was to evaluate the measurement models and not to make conclusions about the population, the representativeness of the sample was not essential. Demographics are given, but there is no assumption of generalizability. In addition, even though no multi-group analysis of males and females occurred during the pilot studies, the author presents stratified demographics to align with the main study presentation.

Table 4

Pilot Study I Demographics

Demographic		Males (n = 41)	Females (n = 46)
Age	*M* (*SD*)	34 (6)	35 (8)
Education level	Less than high school	0	0
	High school graduate	2	2
	Some college (no degree)	1	0
	Associate's degree	1	2
	Bachelor's degree	29	35
	Master's degree	7	7
	Doctorate degree	1	0
	Skipped answers	0	0
Ethnicity	Caucasian	31	44
	African descent	1	0
	Asian descent	7	0
	Hispanic descent	1	1
	Other	0	1
	Skipped answers	1	0
Marital status	Never been married	8	2
	Married	33	43
	Divorced	0	0
	Separated	0	0
	Widow or widower	0	1
	Skipped answers	0	0

There were eight missing values, two of which were by the same participant, though they occurred in different constructs. Five missing values were part of reflective constructs, and one was missing from a global indicator. The researcher imputed them by

examining the responses of associated indicators and using the inference method (see Hair et al., 2010). There were two missing formative values, which were replaced with the middle value of the scale (3-undecided).

Smart PLS4® was used to conduct reliability and validity assessments and estimate path coefficients of redundancy models. To evaluate the significance of outer loadings and outer weights and to develop HTMT intervals, the researcher utilized 10,000-sample bootstrapping. Results from pilot study I showed no operational irregularities with the survey instrument. The following sections address reliability and validity problem areas.

Reflective Indicator Reliability

Several indicators showed significant but low outer loadings ($l < .708$), demonstrating low commonality with the construct. Table 5 shows indicators that did not meet the .708 threshold, although none fell below the .4 threshold, which would warrant removal (see Hair et al., 2022).

Table 5

Pilot Study I Indicators Below .708

Indicator	*l*	*p*
PS3	.54	< .01
RN3	.556	.02
SE1	.697	< .01
SE2	.636	< .01
SE3	.569	< .01
SE4	.583	< .01
SE5	.637	< .01
SE6	.451	< .01
SE7	.686	< .01
VO3	.662	< .01

An outer loading less than .708, such as those in Table 5, is evidence that the indicator shares less than 50% of its variance with the construct. In other words, error influences the variance of the Table 5 indicators more than their construct. Such a large error influence may mean participants attribute a meaning to the indicator that differs from that of the co-indicators. In other words, compared to its co-indicators, the indicator measures something different. Researchers should consider removing indicators displaying such characteristics depending on other reliability and validity measures of the construct (Hair et al., 2022).

Internal Consistency Reliability

Three constructs showed low levels of internal consistency for two of the three evaluation methods—Cronbach's alpha (α), reliability coefficient (ρ_a), and composite

reliability (ρ_c). Table 6 shows constructs with internal consistency outside the desired range.

Table 6

Pilot Study I Constructs Not Meeting Internal Consistency Thresholds

Construct	α	ρ_a	ρ_c
Deciding to act	.471	.519	.784
Recognizing the need for intervention	.486	.479	.728
Victim outcome	.483	.483	.744

The three consistency measures are based on the intercorrelations of the indicators (Hair et al., 2022), and displaying low internal consistency (< .7) means the indicators as a group may not be accurately (or consistently) measuring their target construct. Since Cronbach's alpha tends to underestimate internal consistency and composite reliability tends to overestimate it, the reliability coefficient compromises the two (see Chapter III). As a result, the reliability coefficient is used as the primary reliability measure, although all three are presented to provide readers with the most thorough assessment.

Convergent Validity

Three constructs had low convergence with their indicators (measured by AVE). Table 7 shows the constructs indicating low convergent validity.

Table 7

Pilot Study I Constructs Not Meeting AVE Threshold

Construct	*AVE*
Recognizing the need for intervention	.475
Self-efficacy	.431
Victim outcome	.492

AVE is similar to outer loadings, but whereas loadings are specific to each indicator, AVE is an average of all the explained variance (I^2) of the indicators (see Hair et al., 2022). It quantifies how much of the collective variance of the indicators can be attributed to the overarching construct. The low values in Table 7 mean that, on average, the identified constructs explain less than half of the average variance of their indicators.

Adjustments to Reflective Models

The constructs *recognizing the need for intervention* and *victim outcome* showed consistent problems in terms of indicator reliability (see Table 5), internal consistency (see Table 6), and convergent validity (see Table 7). One option for constructs with such problems is to discard either indicators or the entire measurement model. That option was rejected, however, due to the irreversibility that such an action would have during the main study. Instead, the researcher attempted to eliminate potentially misleading or confounding dimensions by altering verbiage. For the construct *recognizing the need for intervention*, the researcher removed the sentence adverb from each indicator to provide a more precise delineation of intent while changing the sentence subject from participant to flight attendant (see Table 8).

Table 8

Alterations to Recognizing the Need for Intervention Indicator Wording

Indicator	Original Wording	Revised Wording
RN1	It is evident to me that the flight attendant urgently needs help.	The flight attendant urgently needs help.
RN2	I believe in this situation, passengers should get involved.	The flight attendant needs others to get involved.
RN3	I believe that this situation is an emergency that requires help from passengers.	The flight attendant needs help from passengers.

For the lower-order construct *victim outcome*, the researcher changed the wording to isolate the victim's reaction to intervening ("…the victim will feel less affected…") instead of the success of the action itself ("…it would make…") (see Table 9).

Table 9

Alterations to Victim Outcome Indicator Wording

Indicator	Original Wording	Revised Wording
VO1	If I tried to physically make the aggressor stop, it would make the victim feel better.	If I try to physically make the aggressor stop, the victim will feel less affected by this incident.
VO2	If I told the aggressor to stop, it would make the victim feel better.	If I tell the aggressor to stop, the victim will feel less affected by this incident.
VO3	If I told others to make the aggressor stop, it would make the victim feel better.	If I tell others to make the aggressor stop, the victim will feel less affected by this incident.

Even though the construct *deciding to act* did not show consistent problems, the poor internal consistency warranted review. Although the measurement was sourced from the literature, a wording review showed potentially confounding dimensions within the indicators. Specifically, the dimensions of *support* and *looking foolish* may have

contributed to the lack of correlation between the two. To explore that possibility, the *deciding to act* indicators were altered to remove the potentially errant dimensions (see Table 10).

Table 10

Alterations to Deciding to Act Indicator Wording

Indicator	Original Wording	Revised Wording
DA1	I would intervene even if I am not sure other passengers would support me.	I will get up and help the flight attendant.
DA2	I would intervene even though I might look foolish.	I will physically go help the flight attendant.

Most indicator loadings of the construct *self-efficacy* showed medium reliability and convergent validity. Typically, this would warrant consideration for discarding the construct. However, due to the irreversibility such action would have on the main study and the measurement's established reliability in existing literature, the researcher retained all *self-efficacy* indicators in their original format. In addition, since only one indicator for *perceived social influence* (PS3) had medium indicator reliability, and the construct performed well in all other assessments, the wording for PS3 was retained in its original format.

Formative Construct Convergent Validity

The sample size for the pilot study (n = 87) was less than the minimum (n = 92), so the analysis included the recognition that there was an increased possibility of a type I or II error. During the formative convergent validity review (redundancy analysis), all three non-HCM formative constructs had mixed performance. The path coefficient

between each and their single-item construct was high ($\beta > .7$), but the level of explained variance was low ($R^2 < .5$) (see Hair et al., 2022) (see Table 11). Such results indicate that while the constructs have strong covariance with their reflectively measured single-items, the unexplained variance ($1 – R^2$) of the single-item construct exceeds the explained variance (R^2) (see Hair et al., 2022). In terms of construct dimensions, the results could mean that the dimensional boundaries defined by the formative indicators differed from those of the reflective item.

Table 11

Path Coefficient and Explained Variance During Redundancy Analysis (Pilot I)

Construct	β	R^2
Out-group social bias	.714	.310
Intervention skills	.806	.468
Nonintervention cost	.761	.342

The objective of the global item is to capture the full, but exclusive, extent of a participant's belief in the presence of all the formative construct dimensions. After review, it appeared the single-item constructs were capturing whether or not any dimensions existed. Using *intervention skills* as an example, a participant with *only* police training responding to the statement "I have received training that can be used to help someone in need." might answer the same as a participant with all five training types since the original statement is rather binary. In other words, participants may have interpreted the statement as referencing whether or not at least one type of training has

been received, thus failing to capture any distinction between only one type and several types received.

Adjustments to Formative Models

As a result of the above suspicion, the three global items were changed to better capture the missing *variety* aspect (see Table 12).

Table 12

Alterations to Wording of Formative Global Indicators

Indicator	Original Wording	Revised Wording
IS6	I have received training that can be used to help someone in need.	I have received many different types of training that can be used to help someone in need.
SB6	Flight attendants have more admirable qualities than most people.	Compared to most people, how admirable overall are flight attendants?
NC5	If I do nothing, this incident could negatively affect me.	If I do nothing, this incident could negatively affect me in many different ways.

Expectation of Positive Outcome

Similar to the other formative constructs, the HCM *expectation of positive outcome* performed poorly regarding explained variance and correlation with its global construct (see Table 13). These results may indicate that the dimensions captured by the global item either differ or are more numerous than the two dimensions contained in the lower-order constructs (*victim outcome* and *situation outcome*) or that one or both of the dimensions in the lower-order constructs are not part of the higher-order construct. In other words, the lower-order constructs may not converge on the concept defined by the higher-order one. Since the alterations already made to *victim outcome* may affect these results, the remaining HCM structure was retained.

Table 13

Path Coefficient and Explained Variance of Expectation of Positive Outcome Redundancy Analysis (Pilot I)

Construct	β	R^2
Expectation of positive outcome	.54	.321

Pilot Study II

To evaluate the efficacy of the above changes, a second pilot study was conducted utilizing 104 new participants. Internal MTurk restrictions excluded pilot study I participants from pilot study II. Although the researcher requested 100 participants, an overage of 4% (similar to pilot study I) occurred and was attributed to the same cause. As in pilot study I, all participants who completed the MTurk work submission were compensated.

Utilizing the same methods and criteria from pilot study I, one participant was eliminated due to straight-lining and another 23 due to having a high probability of being non-human. During pilot study II, it became apparent that some participants had spent as little as one second on each item (inattentiveness). The Qualtrics platform does not record individual question duration but does report the total time spent with the survey open. Literature is absent guidelines regarding the minimum expected time per question, so the researcher conducted several iterations of the survey by rapidly reading and randomly answering all survey items. Based on the results, the researcher established an average of three seconds per item (171 total seconds) as the threshold required to keep a participant. Two participants were removed for failing to meet that threshold. A sample of $n = 78$ remained. Demographics of the sample are in Table 14.

Table 14

Pilot Study II Demographics

Demographic		Males (n = 54)	Females (n = 24)
Age	*M* (*SD*)	34 (7)	35 (11)
Education level	Less than high school	0	0
	High school graduate	2	1
	Some college (no degree)	0	1
	Associate's degree	1	0
	Bachelor's degree	33	14
	Master's degree	18	8
	Doctorate degree	0	0
	Skipped answers	0	0
Ethnicity	Caucasian	51	22
	African descent	1	0
	Asian descent	1	1
	Hispanic descent	0	0
	Other	0	1
	Skipped answers	1	0
Marital status	Never been married	5	2
	Married	48	21
	Divorced	0	0
	Separated	0	0
	Widow or widower	0	0
	Skipped answers	1	1

There were 10 cases of missing values. No participants had more than one missing value in a construct. Only one participant had two missing values in total. The researcher replaced the missing values for reflective constructs and global indicators (two

cases) using the same procedure from pilot study I. For formative constructs, missing values were replaced with a value of *3-undecided*, following the technique used in pilot study I.

Reflective Indicator Reliability

The changes to indicators RN3 and VO3 resulted in greatly increased outer loadings for RN3 and minor increases for VO3 (see Table 13). The modified RN3 shared a high level of variance with its construct (66%) (desired), while VO3 still shared less than half (45%).

Table 15

Comparison of Pilot Study I and II Results for Outer Loading and Significance

	Pilot I		Pilot II	
Indicator	*l*	*p*	*l*	*P*
RN3	.556	.02	.815	< .01
VO3	.662	< .01	.674	< .01

Internal Consistency Reliability

The internal consistency for the three affected constructs (see Table 16) showed a considerable positive change in *victim outcome* and minor changes in *deciding to act* and *recognizing the need for intervention*. However, all three still failed to achieve the accepted minimum level for Cronbach's alpha and reliability coefficient. The results mean the *victim outcome* indicators far more accurately and consistently measure their construct than those in pilot study I. However, they still fell slightly below the generally accepted threshold of .7 (see Hair et al., 2022). The pilot study II indicators of *deciding to*

act and *recognizing the need for intervention* were slightly more accurate and consistent in measuring their associated construct but also remained below accepted thresholds. As in pilot study I, the primary statistic used to evaluate internal consistency in the present study is the reliability coefficient (ρ_a), though the author presents all three to optimize understanding of results.

Table 16

Comparison of Pilot Study I and II Results for Internal Consistency

Construct	Pilot I			Pilot II		
	α	ρ_a	ρ_c	α	ρ_a	ρ_c
Deciding to act	.471	.519	.784	.588	.588	.829
Recognizing the need for intervention	.486	.479	.728	.495	.531	.74
Victim outcome	.483	.483	.744	.663	.681	.817

Convergent Validity

Changes to the wording for *recognizing the need for intervention* and *victim outcome* resulted in a large positive change in the convergent validity of the latter and a slight change in the former (see Table 17). With these pilot study II *victim outcome* indicators, over half (60%) of the average indicator variance is explained by the construct (desired). For the indicators of *recognizing the need for intervention*, however, still less than half (49%) of their average variance was explained by their construct.

Table 17

Comparison of Pilot Study I and II Results for Convergent Validity

Construct	AVE	
	Pilot I	Pilot II
Recognizing the need for intervention	.475	.491
Victim outcome	.492	.601

Formative Construct Convergent Validity

After altering the global indicators, only *intervention skills* achieved minimum accepted values for both path coefficients and explained variance (see Table 18). The results mean that a one increment change in the *skills* formative construct results in a .788 increment change in the reflective construct. During redundancy analysis, such a high correlation indicates the two constructs are close in meaning (preferred). The high explained variance ($R^2 > .5$) indicates that over half of the variance of the reflectively-measured construct is explained by the formative one (preferred). In terms of convergent validity, it means the formative indicators converge onto the same concept—the one measured by the single reflective item. The correlation in the *out-group social bias* and *nonintervention cost* models weakened, and the explained variance was either reduced or remained approximately unchanged, indicating a minimization of the preferred characteristics above—the formative and reflective construct meanings further diverged and likely represented different concepts.

Table 18

Comparison of Pilot Study I and II Results for Convergent Validity

Construct	Pilot I		Pilot II	
	β	R^2	β	R^2
Out-group social bias	.714	.310	.462	.193
Intervention skills	.806	.468	.788	.506
Nonintervention cost	.761	.342	.635	.381

Expectation of Positive Outcome

Changes to the lower-order construct *victim outcome* increased correlation—close to the minimum accepted value—during redundancy analysis but had little effect on explained variance (see Table 19). These results mean the lower-order concepts—the victim being less affected and the situation stopping—became closer in meaning to the concept of intervention being a helping strategy. However, the low level of explained variance ($R^2 < .5$) indicates that participants may still perceive the higher-order concept more broadly.

Table 19

Comparison of Pilot Study I and II Results for Expectation of Positive Outcome Convergent Validity

Construct	Pilot I		Pilot II	
	β	R^2	β	R^2
Expectation of positive outcome	.54	.321	.679	.373

Final Changes to the Measurement Models

Recognizing the Need for Intervention. Based on the increase in indicator reliability, internal consistency, and the slight increase in convergent validity, the pilot study II version of indicators was selected for use in the main study.

Victim Outcome. Since the construct improved in all aspects during pilot study II, its version of indicators was selected for the main study.

Deciding to Act. The convergent validity improved slightly in pilot study II. However, removing the unique dimensions resulted in the construct losing its discriminant validity with nearly all other constructs (see Table 20). In other words, the indicators of *deciding to act* in pilot study II have more in common with four other constructs than they do with their own, making *deciding to act* no longer unique. Based on these results and the existence of the original indicators in the literature, those from pilot study I were selected for the main study.

Table 20

Comparison of Pilot Study I and II Results for Deciding to Act Discriminant Validity

Construct	HTMT Ratio	
	Pilot I	Pilot II
Knowing what to do	.821	**1.279**
Perceived social influence	**.866**	**1.037**
Self-efficacy	.653	.514
Situation outcome	.702	**1.004**
Taking responsibility	.831	**1.325**

Note. Bold text indicates values above .85 (undesirable). Comparisons of *recognizing the need for intervention* and *victim outcome* are omitted since the indicators differed between the pilot studies.

Formative Global Indicators. Based on the results of the pilot studies (see Table 18), the researcher selected the pilot study II global indicator for *intervention skills* and the pilot study I version of *out-group social bias* and *nonintervention cost*.

Main Study

Data for the main study was collected in September and October 2023. During the collection period, there were no high-profile flight attendant assault cases or airline security events in the U.S. national news. Two identical surveys—one each for males and females—were created and published simultaneously. Internal MTurk restrictions excluded males from the female survey, females from the male, and all pilot study participants from both.

Data Quality

The initial request was for 750 responses each from males and females, though 785 and 770 responses were recorded, respectively. As in the pilot studies, the slight overage (3.6%) was attributed to respondents completing the survey but failing to identify that they had done so in MTurk. Also, during compensation submission, it became apparent that 40 participants had managed to access both the male-only and female-only surveys (80 total responses). The participants were compensated twice, but all 80 responses were removed. However, Amazon refunded more than was paid, allowing for an additional 54 male and 53 female responses to be collected. Initial data quality analysis indicated the male sample size was inadequate ($n < 550$), so another 100 male surveys were collected. In total, the initial sample size consisted of 1,762 participants. All participants were compensated per IRB requirements.

Data quality analysis was conducted with the aggregate sample to ensure consistent criteria were applied across stratification. The researcher also used the same criteria as in the pilot studies. First, the 40 participants who had managed to access both surveys (80 total responses) were discarded. Using Microsoft Excel in the same manner as the pilot studies, the researcher discarded an additional 15 responses due to straight-lining or near straight-lining (identical responses to at least 56 of the 57 questions). Examining the total duration each participant spent taking the survey revealed that four hundred forty-four participants spent less than 171 seconds taking the survey, which equates to less than three seconds per item. These participants and 89 others who showed a high probability of being non-human (Qualtrics CAPTCHA score < .5) were removed. The final usable stratified sample consisted of 555 males and 579 females.

Data Analysis

Data analysis included reviewing demographics for notable anomalies between males and females, handling missing data, and analyzing measurement models for reliability and validity. Bootstrapping (10,000 samples) was used to evaluate the significance of outer loadings, weights, and path coefficients and create HTMT percentiles.

Demographics. Table 21 shows demographic distributions for age, education level, ethnicity, and marital status. The true population is the flying public, but demographics for such a group are not collected. As a proxy, the sample demographics are compared to the adult U.S. population with the recognition that it may differ from the true target population. The two most prominent differences between males and females were a 3.4% difference in the number of high school graduates and a 2.1% difference in

the number of Caucasians. There were no other noteworthy demographic imbalances between males and females.

Table 21

Main Study Demographics

Demographic		Males (*n* = 555)	%	Females (*n* = 579)	%	[a]U.S. Population %
Age	*M* (*SD*)	35 (9)	-	36 (12)	-	39 (-)
Education level	Less than high school	0	0	1	.2	10
	High school graduate	42	7.6	24	4.1	26
	Some college (no degree)	23	4.1	21	3.6	19
	Associate's degree	22	4	29	5	8.8
	Bachelor's degree	343	62	365	63	22
	Master's degree	122	22	134	23	[b]14
	Doctorate degree	3	.5	5	.9	
	Skipped answers	0	0	0	0	-
Ethnicity	Caucasian	487	88	520	90	62
	African descent	16	2.9	17	2.9	12
	Asian descent	19	3.4	10	1.7	6
	Hispanic descent	17	3.1	17	2.9	-
	Other	8	1.4	5	.86	8
	Skipped answers	5	.9	5	.86	-
Marital status	Never been married	66	12	67	12	34
	Married	466	84	477	82	48
	Divorced	19	3.4	22	3.8	10
	Separated	1	.18	4	.69	1.7
	Widow or widower	1	.18	7	1.2	5
	Skipped answers	2	.36	2	.35	-

[a] Age, education level, and marital status data are from "United States: Populations and people," *2022 American Community Survey* by U.S. Census Bureau, n.d.-a, (https://data.census.gov/profile/United_States?g=010XX00US#populations-and-people). In the public domain. . Ethnicity data is from "United States: Race and ethnicity," *2020 Decennial Census* by U.S. Census Bureau, n.d.-b (https://data.census.gov/profile/United_States?g=010XX00US#populations-and-people). In the public domain.

[b] Census data for master's and doctorate degrees was combined as a postgraduate degree.

Participants with less than a Bachelor's degree are under-represented in the sample, as are those who are unmarried and from ethnic minorities. As a result, compared to the U.S. population, the sample is more educated, with a higher percentage of married and Caucasian individuals. Although such variables are not included in this study, some have been found to be influential in parallel bystander research (see Gaertner & Dovidio, 1977; Gaertner et al., 1982). That fact reduces the ability to fully generalize the results to the target population. It also may mean the results may be more useful in a theoretical sense than a practical one, at least until alternate sampling frames can be employed.

Missing Data. Excluding demographics, there were 107 cases (< .2%) of missing data (see Table 22). Missing data less than 1% is considered trivial (Acuna & Rodriguez, 2004), but still must be addressed. Most missing data (numerically and proportionally) was within the *self-efficacy* construct. For missing data within reflective constructs, the researcher imputed the omissions in the same manner as the pilot studies—with the most likely response based on other construct responses (see Hair et al., 2010). The amount of missing data in formative indicators and global items (IS6, SB6, NC5, and EO1) was extremely small (.17%) so missing data was also imputed in the same manner as the pilot studies (median imputation method). No participants had to be discarded due to excessive

(over 15%) missing data (see Hair et al., 2022). One participant had four missing values (8% of the participant's responses), with the remaining having two or less.

Table 22

Quantity and Percentage of Missing Responses

Construct	Number	%
Self-efficacy	48	.42
Perceived social influence	13	.23
Nonintervention cost	12	.21
Expectation of positive outcome (global item)	2	.18
Out-group social bias	12	.18
Intervention skills	8	.12
Victim outcome	4	.12
Situation outcome	3	.09
Knowing what to do	2	.06
Deciding to act	1	.04
Recognizing the need for intervention	1	.03
Taking responsibility	1	.03

Reliability Assessment of Reflective Constructs.

Indicator Reliability. No indicators fell below the removal level ($l < .4$), although one *perceived social influence* indicator (PS3) and all *self-efficacy* indicators (SE1-SE10) had marginal loadings below .7 (see Hair et al., 2022; see Table 23). Such low values indicate that the indicators share less than half of their variance with their construct. These indicators became candidates for removal, dependent on other reliability and validity measures (see Hair et al., 2022).

Table 23

Outer Loadings and Significance

Indicator	*l*	*p*	Indicator	*l*	*p*
DA1	.885	< .01	SE4	.673	< .01
DA2	.827	< .01	SE5	.618	< .01
KW1	.771	< .01	SE6	.613	< .01
KW2	.839	< .01	SE7	.621	< .01
KW3	.801	< .01	SE8	.648	< .01
PS1	.777	< .01	SE9	.596	< .01
PS2	.751	< .01	SE10	.659	< .01
PS3	.493	< .01	SO1	.807	< .01
PS4	.834	< .01	SO2	.855	< .01
PS5	.836	< .01	SO3	.844	< .01
RN1	.736	< .01	TR1	.822	< .01
RN2	.745	< .01	TR2	.807	< .01
RN3	.788	< .01	TR3	.782	< .01
SE1	.693	< .01	VO1	.815	< .01
SE2	.627	< .01	VO2	.857	< .01
SE3	.661	< .01	VO3	.854	< .01

Internal Consistency. Two constructs, *recognizing the need for intervention* and *deciding to act*, scored slightly below .7 for both Cronbach's alpha and reliability coefficient (see Table 24). The results mean the indicators of these constructs were somewhat less reliable (or consistent) in measuring their common dimension. In other words, there were some inaccuracies in the individual indicator variance compared to the combined variance of all the indicators. Readers can conceptualize this as meaning the indicators did not vary as expected compared to their co-indicators.

A value of .7 is widely accepted as the threshold for internal consistency (Taber, 2018; see Hair et al., 2022). Other researchers apply a more liberal threshold of .6 (Van Griethuijsen et al., 2014) or even .5 (see Hair et al., 2010). The measurement values and differing standards created the need to decide whether the two constructs' internal consistency was adequate for retention or inadequate enough to warrant removal. The two constructs were retained based on their exceeding literature-based minimum thresholds (see Van Griethuijsen et al., 2014; see Hair et al., 2010) and the need to maintain reasonable integrity of the core intervention model steps.

Table 24

Internal Consistency Measures of Reflective Constructs

Construct	α	ρ_a	ρ_c
Recognizing the need for intervention	.628	.632	.801
Taking responsibility	.726	.726	.846
Knowing what to do	.727	.733	.846
Deciding to act	.639	.654	.846
Situation outcome	.783	.784	.874
Victim outcome	.795	.796	.88
Self-efficacy	.842	.848	.875
Perceived social influence	.795	.82	.861

Validity Assessment of Reflective Constructs.

Convergent Validity. All constructs except *self-efficacy* showed adequate convergent validity (AVE > .5), so except for *self-efficacy*, all constructs, on average,

explained over half of the variance of their indicators, meaning the indicators indeed converged on the same concept.

Table 25

Average Variance Extracted

Construct	AVE
Deciding to act	.734
Victim outcome	.709
Situation outcome	.698
Knowing what to do	.647
Taking responsibility	.646
Recognizing the need for intervention	.573
Perceived social influence	.561
Self-efficacy	.412

Discriminant Validity. Smart PLS software does not support the technique of using fit indices to evaluate discriminant validity (see Chapter III). However, it does support the Fornell-Larcker criteria and the HTMT methods. According to the Fornell-Larcker criterion, all constructs showed adequate discriminant validity (see Table 26). Since some researchers believe the Fornell-Larcker method tends to under-identify discriminant validity issues (see Hair et al., 2022), the researcher also analyzed discriminant validity via the HTMT method.

Table 26

Fornell-Larcker Measure of Discriminant Validity

	DA	KW	PS	RN	SE	TR	SO	VO
DA	.856							
KW	.565	.804						
PS	.558	.575	.749					
RN	.397	.359	.325	.757				
SE	.328	.430	.312	.400	.642			
TR	.620	.686	.644	.467	.339	.804		
SO	.474	.638	.667	.229	.357	.605	.835	
VO	.432	.556	.662	.273	.321	.57	.709	.842

The HTMT method showed that the construct *taking responsibility* (TR) was not measuring a concept distinct from either *deciding to act* (DA) or *knowing what to do* (KW) (see Table 27). Also, six ratios (DA-KW, KW-SO, PS-TR, PS-SO, PS-VO, and SO-VO) showed a bootstrap upper 95% bound greater than .85, meaning that there is a less than 95% certainty that the actual HTMT value in the population is less than the maximum HTMT value (.85 for dissimilar constructs, .9 for theoretically similar constructs) based on bootstrap sampling. Considering the results from both methods, discriminate validity is achieved for all constructs except *taking responsibility*.

Table 27

Heterotrait-Monotrait Ratio and Upper 95% Bound

	DA	KW	PS	RN	SE	TR	SO
KW	.819 (.882)						
PS	.779 (.845)	.749 (.807)					
RN	.623 (.705)	.531 (.616)	.482 (.563)				
SE	.437 (.513)	.535 (.603)	.384 (.453)	.562 (.647)			
TR	.904 (.968)	.941 (.988)	.845 (.894)	.687 (.76)	.423 (.497)		
SO	.664 (.729)	.84 (.893)	.829 (.87)	.325 (.402)	.417 (.485)	.802 (.85)	
VO	.6 (.67)	.728 (.787)	.819 (.864)	.384 (.466)	.373 (.439)	.75 (.808)	.898 (.939)

Note. Since SO and VO are lower-order constructs of the same HCM, they are more theoretically similar than other unrelated constructs. As a result, their HTMT ratio limit is .9 instead of .85.

Validity Assessment of Formative Constructs.

Convergent Validity. Redundancy analysis of the construct *intervention skills* showed acceptable path coefficient and explained variance (see Table 29). *Out-group social bias* and *nonintervention cost* also had good path coefficients, but low explained variance. Such results likely indicate that the global construct encompasses more dimensions than the formative indicators represent which compromises the understanding of their relationships. For instance, participants understood and imagined the construct *out-group social bias* to be much broader and less constrained than what was described by the five indicators. That reduces the understanding of the relationship *out-group social bias* has with its target construct, since the apparent true meaning participants applied to the *social bias* construct is unclear. The relationship between *nonintervention cost* and its target construct suffers in a similar fashion, though not as severely.

Table 28

Path Coefficient and Explained Variance During Redundancy Analysis (Main Study)

Construct	β	R^2
Out-group social bias	.885	.296
Intervention skills	.881	.582
Nonintervention cost	.840	.411

Content Validity. Since *nonintervention cost* was an original outer model (not from the literature), eight SMEs evaluated it for content validity. The SMEs were provided a link (https://vimeo.com/852179353) to a researcher-developed video defining content validity, explaining their role, and giving instructions for an online Qualtrics survey explicitly designed for the analysis (see Figure D1). SMEs were free to contact the researcher with questions, though none did. The results were analyzed using the modified kappa index described in Chapter III.

One indicator (NC2) had only fair content validity, while the remaining three scored excellent (see Table 29). The results mean that according to SME analysis, NC1, NC3, and NC4 clearly belong to the construct, but NC2 may very well be a minor dimension at best.

Table 29

Content Validity of Nonintervention Cost

Indicator	*k**
NC1	.87
NC2	.52
NC3	.87
NC4	1

Additional Measurement Model Assessments. In addition to the above evaluations, the collinearity, significance, and relevance of formative indicators and the HCM were also assessed.

Collinearity of Formative Indicators. All formative indicators showed low collinearity (VIF < 3) (see Hair et al., 2022) (desired) with their associated indicators (see Table 30). Low collinearity is evidence that each indicator measures a unique dimension of its construct and has minimal dimensional overlap with its co-indicators.

Table 30

VIF of Formative Indicators

Indicator	VIF	Indicator	VIF
IS1	2.351	NC3	1.384
IS2	2.209	NC4	1.320
IS3	1.817	SB1	1.283
IS4	1.944	SB2	1.221
IS5	2.274	SB3	1.294
NC1	1.608	SB4	1.278
NC2	1.503	SB5	1.186

Significance and Relevance of Formative Indicators. Two indicators (NC2 and SB5) had nonsignificant outer weights and low outer loadings ($l < .5$) (see Hair et al., 2022; see Table 31). The combination means that not only do the indicators contribute little relative to their co-indicators in forming the construct (w), but their absolute contribution (l) is also minimal, albeit statistically significant. As a result, they both become candidates for removal.

Table 31

Outer Weights, Loadings, and Significance of Formative Indicators

Indicator	*w*	*p*	*l*	*p*
IS1	0.379	< .01	0.877	< .01
IS2	0.187	< .01	0.789	< .01
IS3	0.155	< .01	0.697	< .01
IS4	0.304	< .01	0.835	< .01
IS5	0.193	< .01	0.816	< .01
NC1	0.244	< .01	0.663	< .01
NC2	-0.011	.454	0.489	< .01
NC3	0.535	< .01	0.832	< .01
NC4	0.492	< .01	0.809	< .01
SB1	0.514	< .01	0.815	< .01
SB2	0.131	< .01	0.430	< .01
SB3	0.290	< .01	0.672	< .01
SB4	0.430	< .01	0.764	< .01
SB5	0.005	.459	0.305	< .01

Hierarchical Component Model. The collinearity of the lower-order constructs (as indicators of the higher-order construct) was low (desired) (see Table 32), meaning

each is unique in their measured concept, or there is little overlap in what they represent. Also, the outer weights of the lower-order constructs (as indicators) were both significant, and their outer loadings showed a high contribution ($l > .5$) in forming the higher-order construct (see Table 33). The combination means both significantly contributed to forming the higher-order construct in absolute terms (l) and relative to each other (w).

Table 32

VIF of Lower-Order Constructs as HCM Indicators

Indicator	VIF
SO	2.012
VO	2.012

Table 33

Outer Weights, Loadings, and Significance of Lower-Order Constructs as HCM Indicators

Indicator	*w*	*p*	*l*	*p*
SO	.69	< .01	.963	< .01
VO	.384	< .01	.874	< .01

The HCM performed poorly during redundancy analysis (to assess convergent validity). The path coefficient and explained variance were less than desired (see Table 34). The results indicate that unobserved variables attributing to the variance of the global indicator have more influence than the two lower-order constructs. In other words, the global item likely contains dimensions more numerous than the two lower-order constructs. In terms of measurement theory, the operationalization of *expectation of*

positive outcome (measured by its global item) is too broad. It should be narrowed to include only the victim and situation dimensions.

Table 34

Path Coefficient and Explained Variance of Expectation of Positive Outcome Redundancy Analysis (Main Study)

Construct	β	R^2
Expectation of positive outcome	.493	.249

Modifications to the Structural Model

Due to the indicator NC2 having nonsignificant outer weights, low outer loading, and only fair content validity, the researcher removed it from the model. The researcher also removed the indicator SB5 due to its nonsignificant outer weight and low outer loading. Both of these indicators contributed little to forming their construct. Doing so improves the validity of the *nonintervention cost* and *out-group social bias* outer models (and ultimately the structural model) by removing noncontributing indicators which would, if retained, inflate the estimation error.

The construct *taking responsibility* and its indicators performed well in all aspects except the HTMT method of discriminant validity. HTMT ratios so close to a value of "one" mean that the indicators of *taking responsibility* share almost as much correlation with the indicators of *deciding to act* and *knowing what to do* as they do with each other. Such a high correlation shows that the *taking responsibility* indicators may not distinctly measure their construct. There are three options for managing such a high correlation (Hair et al., 2022):

a. Eliminate indicators having weak correlations within the construct.
b. Eliminate indicators having strong correlations outside the construct.
c. Split or combine constructs with the possibility of creating an HCM.
d. Discard the measurement model.

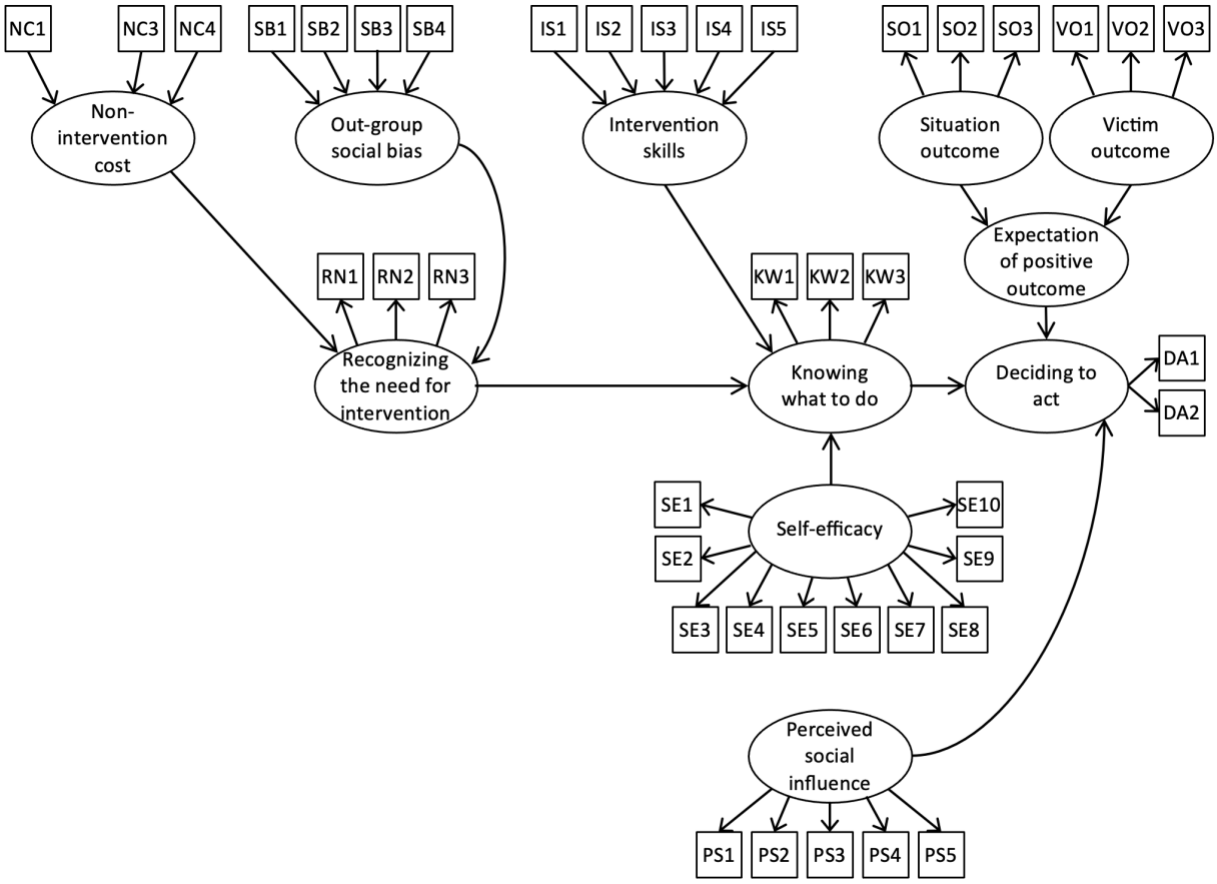
However, researchers should be cautious about selecting options *a*, *b*, or *c* solely for statistical reasons (Hair et al., 2022). These options should only be utilized if doing so is supported by literature, and the researcher is assured that eliminating indicators does not omit relevant dimensions of the construct (via prior content validity analysis) (Hair et al., 2022). Literature does not support combing steps of the BIM (see Aydemir & Gleibs, 2021; see Nickerson et al., 2014), and only by eliminating two of the three indicators can HTMT ratios be established to indicate discriminant validity. However, the latter action narrows the construct dimension considerably. Based on these considerations, the researcher removed *taking responsibility* from the research model (option *d*.), established *out-group social bias* as a predictor of *recognizing the need for intervention*, and the latter as a predictor of *deciding to act*. The literature does not include *out-group social bias* predicting *recognizing the need for intervention*. However, since the outer model performed well, the researcher retained it in the inner model. In a somewhat exploratory step, the researcher evaluated the path coefficients between *out-group social bias* and the remaining steps of the BIM and found that it correlated with *recognizing the need for intervention* better than the other two steps.

Although *self-efficacy* had marginal outer loadings and AVE, it performed well regarding reliability and discriminate validity. Based on these results, it was retained in the model (see Hair et al., 2022). The indicator PS3 also had a marginal outer loading

(see Table 23). However, its associated construct (*perceived social influence*) performed well in all other reliability and validity aspects, so the researcher retained it. The initial research model (see Figure 6) was modified to that shown in Figure 8.

Figure 8

The Modified Research Model



Assessment of the Structural Equation Model

Assessing the structural model entailed initially identifying collinearity issues with co-predictors, followed by a multi-group analysis, which involves evaluating the level of invariance between male and female groups. The relationships, explanatory

power, and predictive power were also examined within the confines of the invariance testing.

Collinearity of the Structural Model

Like the collinearity assessment of formative indicators, the researcher assessed co-predictors of endogenous constructs for collinearity. Table 35 shows that all predictor groupings had low collinearity (VIF < 3) (see Hair et al., 2022), meaning there was little redundancy in the variance attributed to predictors. Too much collinearity within predictor groupings could result in erroneous path coefficients due to artificially high correlations during model estimation (see Hair et al., 2022).

Table 35

Collinearity of Construct Predictors

Construct	Predictor	VIF
Recognizing the need for intervention	Nonintervention cost	1.245
	Out-group social bias	1.245
Knowing what to do	Intervention skills	1.1
	Recognizing the need for intervention	1.197
	Self-efficacy	1.272
Deciding to act	Knowing what to do	1.825
	Expectation of positive outcome	2.496
	Perceived social influence	2.128

Multi-Group Analysis

Since the permutation method is sensitive to different group sizes (Hair et al., 2022), the researcher numbered all female participants, used Microsoft Excel to generate

24 non-repeating random numbers from one to 579, and removed those participants from the female group. The two groups were used to perform invariance testing and a permutation MGA.

Invariance Testing. Before combining or comparing male and female groups, an evaluation of invariance must occur. Per the MICOM procedure, establishing configural and compositional invariance (partial measurement invariance) is the minimum required to compare groups (Hair et al., 2022). Establishing equality of means and variances (full measurement invariance) allows the two groups to be combined, thereby increasing statistical power and generalizability (Hair et al., 2022).

Configural Invariance. The researcher specified identical outer models for each group and treated data quality issues with the aggregate dataset, ensuring identical elimination criteria and missing value treatment across groups. In addition, Smart PLS4 estimates models for each group using full dataset criteria, meaning users apply criteria for the entire dataset prior to group separation. Doing so ensured that the settings in Table 36 applied to both groups. Since outer models, data treatment, and settings were the same for both male and female groups, configural invariance was established for all constructs.

Table 36

Model Estimation Settings

Criteria	Setting
Initial weights	1
Maximum number of iterations	3,000
Stop criterion	10^{-7}
Weighting scheme	Path
Outer weighting mode[a]	Mode A (reflective) Mode B (formative)

[a] Mode A/B refers to the method used to estimate outer weights and loadings. Mode A uses correlation, Mode B uses linear regression (Hair et al., 2022).

Compositional Invariance. Compositional invariance establishes whether or not the two groups conceptualize the constructs the same. Specifically, it identifies statistically different composite scores between groups (not desired) (Hair et al., 2022). Compositional invariance is violated if groups attribute different meanings or intrinsic values to one or more indicators. Such an occurrence would manifest in different outer weights or loadings and, thus, different composite scores. Compositional invariance is established if the difference in composite scores across groups is statistically insignificant ($p > .05$) (Hair et al., 2022).

The results in Table 37 show that all measurement models except *knowing what to do* and *self-efficacy* display compositional invariance. Statistically, the results mean that males and females weighted the observed variables differently in creating the two composite scores. Intrinsically, the results mean that the two groups applied different meanings or worth to the observables. In other words, males and females have different

interpretations or assign different levels of importance to the indicators of *knowing what to do* and *self-efficacy*. The revelation is particularly notable regarding self-efficacy, as it is a well-established measurement model (see Schwarzer, 1999). Equally notable is that some researchers have identified self-efficacy as the crucial element in a bystander knowing what to do (Gini et al., 2008; Pöyhönen et al., 2010), but males and females in the present research viewed both concepts differently.

Table 37

Significance of Composition of Construct Difference Between Males and Females

Measurement model	*p*	Compositionally invariant
Deciding to act	.08	Yes
Intervention skills	.055	Yes
Nonintervention cost	.245	Yes
Out-group social bias	.256	Yes
Perceived social influence	.648	Yes
Recognizing the need for intervention	.152	Yes
Situation outcome	.092	Yes
Victim outcome	.159	Yes
Knowing what to do	< .01	No
Self-efficacy	.001	No

Achieving compositional invariance—and thus partial measurement invariance—allows for meaningful comparisons of male and female groups (Hair et al., 2022), but only for those model relationships exclusive of *knowing what to do* and s*elf-efficacy*. For model relationships involving the latter two, readers must analyze the results for males and females separately and without comparison.

Equality of Composite Means and Variances. To establish full measurement invariance, mean composite scores for the two groups must be equal, and the variances must be the same (or at least not statistically different; Hair et al., 2022). Graphically, the distribution curves must match. Such similarity would mean that the two groups conceptualize the constructs the same and that the variation within each group is the same. Equality of means and variances is established if the difference between groups for both criteria is statistically *insignificant* ($p > .05$) (Hair et al., 2022).

Table 38 shows the results of the equality assessment. Of the compositionally invariant measurement models, only three are fully invariant. The results mean that the samples for males and females can be combined, though only for portions of the structural model that exclusively contain the three constructs. Group data cannot be combined for the remaining structural model, though meaningful comparisons between males and females are still possible.

Table 38

Significance of Composite Mean and Variance Difference Between Males and Females

Measurement model	*p* (*M*)	*p* (*Var*)	Fully measurement invariant
Nonintervention cost	.146	.067	Yes
Out-group social bias	.074	.131	Yes
Recognizing the need for intervention	.576	.275	Yes
Deciding to act	< .01	.758	No
Intervention skills	.002	.514	No
Perceived social influence	< .01	.579	No
Situation outcome	< .01	.008	No
Victim outcome	< .01	.224	No

Note. Only compositionally invariant measurement models are shown.

Invariance testing showed mixed results. Achieving full measurement invariance for *nonintervention cost*, *out-group social bias*, and *recognizing the need for intervention* means the entire dataset (n = 1,134) is available during estimation, but only for the two associated relationships (see Figure 8) and only if the two groups are homogeneous in terms of path coefficient and explained variance. The remaining partially invariant measurement models (see Table 37) allow for comparing male and female path coefficients and explained variance, but only for the two relationships involving *expectation of positive outcome* and *perceived social influence* influencing *deciding to act* (see Figure 8). The remaining structural relationships and explained variances for males and females must be analyzed separately.

Relationships

All relationships in the structural model except *expectation of positive outcome* influencing *deciding to act* (EP—DA) were positive and statistically significant at the $p <$ .01 level (see Table 39). However, the sample size is not large enough to reliably detect minimal relationships ($\beta <$ .16) (see Chapter III), so the relationship between *self-efficacy* (SE) and *knowing what to do* (KW) for the female group is susceptible to being a type I error. All other significant relationships were greater than the $\beta =$ .16 level.

Table 39

Significance, t-Value, and Confidence Intervals of Structural Relationships

Predictor	Target	Male				Female				Diff (*p*)
		β_S [β_U]	*p*	*t*	*CI*	β_S [β_U]	*p*	*t*	*CI*	
SB	RN	.191 [.207]	< .01	4.1	.09, .273	.275 [.292]	< .01	5.7	.173, .363	.223
SB	RN	(results for male and female groups combined)				.237 [.259]	< .01	6.9	.174, .307	-
NC	RN	.396 [.299]	< .01	8.4	.3, .486	.218 [.168]	< .01	4.8	.12, .298	.012
EP	DA	.043 [.049]	.489	.7	-.088, .155	-.024 [-.024]	.736	.3	-.175, .097	.469
PS	DA	.308 [.33]	< .01	4.9	.188, .437	.376 [.388]	< .01	5.7	.255, .512	.466
KW	DA	.42 [.498]	< .01	7.4	.31, .531	.297 [.308]	< .01	4.9	.18, .417	[a]-
IS	KW	.456 [.322]	< .01	12	.378, .522	.632 [.481]	< .01	20	.569, .69	[a]-
RN	KW	.234 [.27]	< .01	4.6	.139, .338	.174 [.24]	< .01	4.4	.097, .254	[a]-
SE	KW	.232 [.314]	< .01	4.5	.133, .329	.144 [.229]	< .01	3.2	.049, .228	[a]-

Note. β_S = standardized, β_U = unstandardized.
[a] Male and female groups cannot be compared due to failing to achieve compositional invariance.

Even though researchers more commonly report standardized path coefficients (β_S), unstandardized coefficients (β_U) are also helpful and often easier to interpret—assuming scales are consistent across measurement models (as in this research). In the present results, a one-unit change in the Likert scale of a predictor results in a β_U-unit change in the target. For example, an increase of "one" in the Likert scale for female

intervention skills correlates with a .481 Likert scale change in *knowing what to do*. Table 39 provides both coefficients to allow readers to fully understand the relationships.

The nonsignificant (and minimal) path coefficient between e*xpectation of positive outcome* (EP) and *deciding to act* (DA) means that males' and females' decisions to help the flight attendant are not influenced by how successful they expect to be, at least in terms of stopping the altercation and reducing its impact on the victim. Such results may indicate the presence of influencers outside the assault triad, such as a moral obligation or social expectation.

Among the four model relationships available for male-female comparison, only one showed a significant difference between the two groups—*nonintervention cost* (NC) influencing *recognizing the need for intervention* (RN). Such a difference may indicate that males are more influenced by how the event will impact them personally, meaning they more closely align with the idea of the self-centered bystander. The fact that the remaining three comparable relationships showed no significant difference between males and females is noteworthy. It may indicate fewer gender-based social expectations of behavior among air travelers. Both males and females may see behavioral expectations as equalized, driven more by their status as a passenger and less by their gender roles.

The researcher used 10,000-sample bootstrapping to evaluate significance using three measures—significance level (*p*), *t*-value, and confidence intervals (*CI*), though all three are comparable regarding interpretive meaning. A *t*-value of 2.57 indicates a significance level of .01, with larger values indicating even greater significance (Hair et al., 2022; *t*-value and *p* are inversely related). The confidence interval indicates the range within which 95% of bootstrap sample path coefficients existed. An interval *not*

including zero means there is a 95% certainty that the true path coefficient is also not zero (Hair et al., 2022). The interval is also helpful in evaluating the stability of the path coefficient (Hair et al., 2022). A narrow interval (IS to KW) indicates a more statistically stable relationship than a wide one (PS to DA). Table 39 shows that all significant levels, *t*-values, and *CI*s agree regarding path coefficient significance. The table presents all three to provide optimum data clarity.

Explanatory Power

Table 40 lists each endogenous construct's coefficient of determination (R^2). The table also provides the adjusted coefficient ($\bar{R}^2$), representing the explained variance adjusted for the number of predictors (Hair et al., 2010). Since R^2 represents the sum of explained variance (including type I error variance), it can often increase simply by adding variables (Hair et al., 2010; Hair et al., 2022). That fact can lead to misleading results. $\bar{R}^2$, however, is independent of the number of predictors and better represents the actual amount of explained variance (Hair et al., 2010). To maximize understanding, Table 40 lists both statistics for all endogenous constructs and the difference between male and female groups. Both values in Table 40 directly equate to the percentage of variance explained by a construct's predictors. For example, for the female group, 52% of the variance in *knowing what to do* can be attributed to its exogenous constructs (*self-efficacy*, *intervention skills*, and *recognizing the need for intervention*). The remaining 48% of variance is due to unknown variables (error). The significant difference between males and females in *recognizing the need for intervention* is evidence of the different influence variables can have on the two groups.

Table 40

Coefficient and Adjusted Coefficient of Determination of Endogenous Constructs

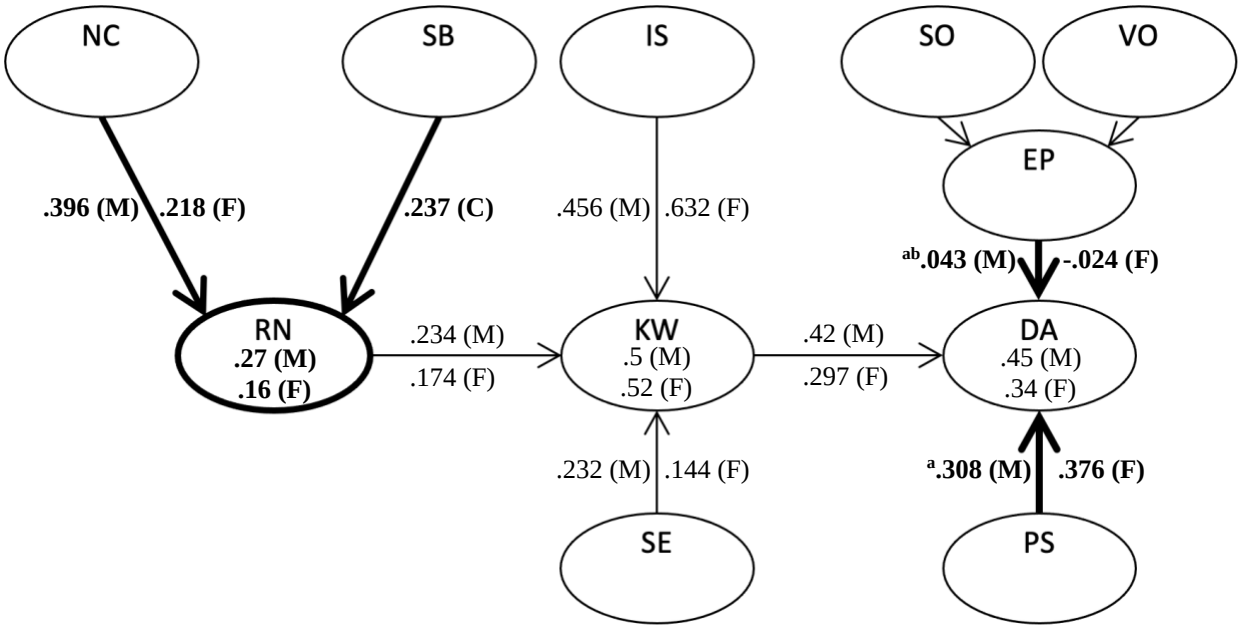
Endogenous construct	Male R^2 ($\bar{R}^2$)	Female R^2 ($\bar{R}^2$)	Diff (*p*)
Recognizing the need for intervention	.272 (.27)	.163 (.16)	.035
Deciding to act	.454 (.451)	.339 (.335)	[a]Cannot compare
Knowing what to do	.497 (.495)	.525 (.522)	[a]Cannot compare

[a] *Knowing what to do* did not achieve partial measurement invariance, meaning *deciding to act* can also not be compared, since *knowing what to do* also acts as a predictor.

Figure 9 illustrates the modified research model with associated path coefficients and explained variance.

Figure 9

Path Coefficients and Explained Variance for (M)ales, (F)emales, and (C)ombined (Where Appropriate)



Note. Bold text indicates male and female groups can be compared. Explained variance is $\bar{R}^2$.

[a] Groups are not statistically different.

[b] EP to DA is nonsignificant.

Predictive Power

Evaluating predictive power is much more relevant for predictive models than it is for theoretical ones (such as the present model). However, some consider it to be an inherent part of PLS analysis. As such, it is presented here for informative purposes only, since it does not add value to answering either research question.

The sample size for males was only slightly more than the minimum (n = 555, n_{min} = 550) which severely limits the ability of $PLS_{predict}$ to return valid results. Typically, a sample of minimum size is extracted from the research sample and used to predict the

values in the remaining portion of the sample. The process is repeated several times to ensure all participant data is used as both predictor and test samples (see Hair et al., 2022). However, each iteration must have a sample size of $n = 550$ as predictor data (see Hair et al., 2022). The requirement means that not only is the sample being predicted incredibly small ($n = 5$), but all iterations of the predictor sample vary by only five participants. As a result, even though the predictive power results are given, readers should be aware of the diminished ability to determine the true out-of-sample predictive power.

Utilizing the maximum number of folds (100) and 99 repetitions yielded the results in Table 41. Small MAE values are desired, representing the mean difference between the predicted and observed (true) values (Hair et al., 2022). In this case, MAE represents the average difference between the predicted Likert scale values and the true values selected by the participants. MAE is always positive, so readers should not interpret the statistic as meaning the true value is greater, but only that it is different by the MAE value (greater or less). Small RMSE values are also desired, but RMSE gives more importance to large errors (Hair et al., 2022). Such errors are weighted more, meaning errors farther from the predicted values contribute more to RMSE than those close. MAE does not make such a distinction. Because of this difference, RMSE is the preferred option (Hair et al., 2022). Table 41 presents both to provide full interpretive options.

Table 41

Male and Female Group MAE and RMSE Values for Endogenous Constructs

	Male		Female	
Endogenous construct	MAE	RMSE	MAE	RMSE
Recognizing the need for intervention	.634	.868	.686	.932
Knowing what to do	.512	.735	.519	.720
Deciding to act	.579	.796	.642	.855

Many readers find MAE and RMSE values challenging to qualify since, without any comparable value or standard, the level of acceptable error becomes highly dependent on the model context. An alternate method to assess predictive power is to compare MAE and RMSE to an LM benchmark (Hair et al., 2022). Table 42 lists each endogenous indicator's MAE and RMSE values and their associated LM values. MAE and RMSE values less than their LM value are preferred, and more preferred values indicate better predictive power (Hair et al., 2022). Because approximately half of the MAE and RMSE values were less than their associated LM values, the model achieved medium predictability (see Hair et al., 2022) for both groups regardless of which technique (MAE or RMSE) is used. The female group, however, had a slightly better performance.

Table 42

MAE, RMSE, and Their LM Benchmark Values for Endogenous Indicators

	Male				Female			
	MAE	LM_{MAE}	RMSE	LM_{RMSE}	MAE	LM_{MAE}	RMSE	LM_{RMSE}
DA1	**0.641**	0.643	0.823	0.821	**0.645**	0.665	**0.853**	0.870
DA2	0.740	0.738	**0.932**	0.939	0.796	0.796	**1.019**	1.024
KW1	0.631	0.630	0.809	0.805	0.665	0.659	**0.842**	0.844
KW2	**0.637**	0.654	**0.807**	0.809	0.602	0.598	0.788	0.780
KW3	**0.642**	0.658	**0.833**	0.868	0.669	0.665	0.853	0.839
RN1	0.623	0.606	0.806	0.792	**0.588**	0.591	0.750	0.743
RN2	0.622	0.611	0.800	0.795	**0.585**	0.596	0.785	0.785
RN3	0.607	0.595	**0.778**	0.779	**0.590**	0.608	**0.773**	0.787

Note. Bold text indicates preferred MAE/RMSE values (< LM).

Hypothesis Testing

Since the construct *taking responsibility* was removed for failing to meet discriminant validity thresholds, there is no statistical way to assess its hypothesized relationships (H_1, H_2, H_5, and H_8). However, since the hypotheses are statements of fact, an inability to justify the fact makes them unsupported. Also, H_1, H_2, H_5, and H_8 sub-hypotheses could not be assessed since they require a statistical basis. Lastly, four sub-hypotheses ($H_{3\text{-}1}$, $H_{6\text{-}1}$, $H_{7\text{-}1}$, and $H_{12\text{-}1}$) could also not be assessed since the two groups were not at least partially invariant. Table 43 lists all hypotheses, the associated statistical justification, and whether each is supported. When the model was modified, it created two newly-hypothesized relationships (H_{11} and H_{12}), which are also included in the assessment.

Table 43

Assessment of Hypotheses

	Hypothesis	β_M (p) β_F (p)	Supported
H_1	Recognizing the need for intervention significantly positively influences taking responsibility.	Removed from model	No
$H_{1\text{-}1}$	There is a significant difference between males and females in the relationship between recognizing the need for intervention and taking responsibility.	Cannot statistically assess	
H_2	Taking responsibility significantly positively influences knowing what to do.	Removed from model	No
$H_{2\text{-}1}$	There is a significant difference between males and females in the relationship between taking responsibility and knowing what to do.	Cannot statistically assess	
H_3	Knowing what to do significantly positively influences deciding to act.	.42 (< .01) .297 (< .01)	Yes
$H_{3\text{-}1}$	There is a significant difference between males and females in the relationship between knowing what to do and deciding to act.	Cannot statistically assess	
H_4	Nonintervention cost significantly positively influences recognizing the need for intervention.	.396 (< .01) .218 (< .01)	Yes
$H_{4\text{-}1}$	There is a significant difference between males and females in the relationship between nonintervention cost and recognizing the need for intervention.	p = .012	Yes
H_5	Out-group social bias significantly positively influences taking responsibility.	Removed from model	No
$H_{5\text{-}1}$	There is a significant difference between males and females in the relationship between out-group social bias and taking responsibility.	Cannot statistically assess	
H_6	Intervention skills significantly positively influences knowing what to do.	.456 (< .01) .632 (< .01)	Yes
$H_{6\text{-}1}$	There is a significant difference between males and females in the relationship between Intervention skills and knowing what to do.	Cannot statistically assess	
H_7	Self-efficacy significantly positively influences knowing what to do.	.232 (< .01) .144 (< .01)	Yes
$H_{7\text{-}1}$	There is a significant difference between males and females in the relationship between self-efficacy and knowing what to do.	Cannot statistically assess	
H_8	Perceived social influence significantly positively influences taking responsibility.	Removed from model	No

	Hypothesis	β_M (p) β_F (p)	Supported
$H_{8\text{-}1}$	There is a significant difference between males and females in the relationship between social influence and taking responsibility.	Cannot statistically assess	
H_9	Perceived social influence significantly positively influences deciding to act.	.308 (< .01) .376 (< .01)	Yes
$H_{9\text{-}1}$	There is a significant difference between males and females in the relationship between social influence and deciding to act.	p = .466	No
H_{10}	Expectation of positive outcome significantly positively influences deciding to act.	.043 (.489) -.024 (.736)	No
$H_{10\text{-}1}$	There is a significant difference between males and females in the relationship between expectation of positive outcome and deciding to act.	p = .469	No
[a]H_{11}	Out-group social bias significantly positively influences recognizing the need for intervention.	.191 (< .01) .275 (< .01)	Yes
[a]$H_{11\text{-}1}$	There is a significant difference between males and females in the relationship between out-group social bias recognizing the need for intervention	p = .223	No
[a]H_{12}	Recognizing the need for intervention significantly positively influences knowing what to do.	.234 (< .01) .174 (< .01)	Yes
[a]$H_{12\text{-}1}$	There is a significant difference between males and females in the relationship between recognizing the need for intervention and knowing what to do.	Cannot statistically assess	

[a] Hypotheses 11 and 12 are added due to the modification of the research model (see Figure 8).

Summary

Chapter IV explained how the researcher conducted both pilot and main studies. The details included the data treatment, assessing the outer models' validity and reliability, and reasons for modifying the researcher model. The chapter provided explanations of the inner-model relationships, explanatory power, and predictive power. The chapter closed with an itemized assessment of the hypotheses.

Chapter V: Discussion, Conclusions, and Recommendations

The final chapter introduces discussions on the quality of the research data and the difficulty encountered with formative content validity. The discussion also explores many of the measurement models and all structural relationships. The research questions and recommendations from participants and SMEs are also presented. Conclusions based on inner and outer models are given, followed by an evaluation of the applicability of the BIM. The chapter closes with practical recommendations for the commercial aviation industry and theoretical recommendations for the research community.

Discussions

Data Quality

Twenty-seven percent of the initial sample failed to meet quality standards, with the vast majority being inattentive participants. The MTurk recommendation (for high-quality work) of using workers with a high completion history may only be effective for work that allows the requester to conduct a quality assessment before reward allocation. For surveys, particularly academic surveys, requiring a high completion history may mean requesters are merely attracting workers who are savvy enough to know that the reward will be forthcoming regardless of quality. Better quality might be achieved by restricting workers to those with *less than* a given threshold, ensuring a more casual, less sophisticated, but more attentive worker. Research into such a strategy may be warranted.

Redundancy Analysis

The redundancy method proved difficult for evaluating the convergent validity of formative constructs. The problem arises from using an unevaluated single-item reflective construct to assess an unevaluated formative one. Such difficulty may be why

researchers have mostly avoided evaluating formative content validity. Poor redundancy analysis results presented the dilemma of identifying which construct was causing the problem—the reflective or the formative. Metaphorically, it is like attempting to hit two targets simultaneously in the dark and only being told that you did not hit both. Determining which direction to adjust becomes challenging, and research is absent corrective techniques when redundancy analysis proves problematic. Sophisticated methods are available to generate single-item reflective measures (see Hair et al., 2022), but successfully doing so would bring into question the need for the formative one. Researchers should continue developing methods to validate formative construct convergent validity and develop adjustment options when redundancy analysis proves problematic.

Measurement Models

Most of the measurement models showed no remarkable characteristics, but six are worthy of discussion. Where applicable, possible explanations for reliability and validity deviations are presented.

Nonintervention Cost. This research took the commonly used notion of a bystander suffering consequences when they intervene (see Allen, 1968; see Fischer & Greitemeyer, 2013; see Latané & Darley, 1970; see Latané & Nida, 1981) and somewhat reversed it by creating a scenario in which the bystander suffers consequences when they fail to intervene. Although rare, the concept can be found in the literature, but not as a measured variable (see Austin, 1979; see Latané & Darley, 1968; see Morgan, 1978). Despite being novel, the measurement model generally performed well and supported the idea that costs to a bystander are not solely from intervening but can be from remaining

passive. Faced with the dilemma, bystanders may undergo a complex evaluation process (see Austin, 1979; see Piliavin et al., 1981), weighing intervention options and their repercussions against those of passivity.

One indicator of the model (the monetary cost of nonintervention) was found to be a nonsignificant contributor to the construct. Participants gave little importance to the idea of monetary loss being a concern when a flight attendant is assaulted. SMEs confirmed this. They also felt the idea was a misplaced addition to the construct. Both results indicate that participants and SMEs likely viewed nonintervention cost in terms of more immediate repercussions, such as getting hurt or having the flight diverted, and less in terms of something that would not be realized until later, such as having to pay for a hotel due to missing a connection. The suspicion is also supported by the strong convergent validity, showing that the participants indeed conceptualized the model as intended, meaning personal negative results (injury, disrupted travel plans, anxiety, and fear) might be realized if they do not act.

Out-Group Social Bias. Similar to *nonintervention cost*, the construct *out-group social bias* reverses the literature-based assumption that bias toward out-group members is universally negative (see Abbott & Cameron, 2014; see Brewer, 2001; see Brewer, 1979; see Doise et al., 1972; see Gardham & Brown, 2001; see Hewstone et al., 2002; see Levine et al., 2005, 2020; see Rabbie & Horwitz, 1969; see Stephenson et al., 1976; see Tajfel, 1979, 1982; see Tajfel et al., 1971; see Wilson & Miller, 1961). In this research, the bias toward the out-group was positive. Of the five dimensions, *well-groomed* was the only one participants felt should be excluded (inferred via outer weights). The dimension was unique, being the only visible quality. The other four dimensions were all

personality-based characteristics (honest, friendly, hard-working, and intelligent). The distinction is evidence that passengers lend value to the flight attendant profession based on its members' personality and not appearance. The omission is noteworthy, as prior research has indicated that the attractiveness of a victim can positively influence males to intervene (Benson et al., 1976). However, Benson's (1976) results may be outdated. The present results may better indicate more current standards of social discernment.

The four remaining indicators strongly influenced how participants viewed flight attendants' overall admirability. However, the four were only a small number of the variables attributing to its makeup (via convergent validity analysis). The fact indicates that participants likely use many other variables to evaluate the admirability of flight attendants. If the above discussion holds, such variables are likely personality-based (humble, generous, or helpful) instead of physical (attractive, fit, or stylish).

Intervention Skills. The strong performance of the *intervention skills* outer model adds to the literature that establishes it as a viable part of intervention research (see Banyard, 2007; see Banyard et al., 2007; see Huston et al., 1981; see Laner et al., 2001; see Potts & Lynch, 2010). Its strong convergent validity indicates that the five skills (police, life-saving, first-aid, self-defense, and medical) are valid dimensions of the construct and comprise the majority of the helping attributes. Despite the diverse skills found in the indicators, the convergent validity adds to Huston et al.'s (1981) findings that the competency to help may not be tied to a specific skill but instead to an ability to handle emergencies in general.

Expectation of Positive Outcome. The stable performance of the HCM indicates the motivations of a victim-centered bystander and situation-centered bystander can be

viewed in terms of the specifics a bystander focuses on in their general expectation of what their intervention will do. Up to this point, the literature has only included the independent parts of the HCM—victim-centric (see Davis, 1994; see DeSmet et al., 2016; see Eisenberg, 1991; see Eisenberg & Fabes, 1998; see Endresen & Olweus, 2001; see Gini et al., 2007, 2008; see Hoffman, 2001; see Philpot et al., 2020; see Pöyhönen et al., 2012; see Rudolph et al., 2004; see Weiner, 1980) and situation-centric (see Dillon & Bushman, 2015; see Pöyhönen et al., 2012). Other research is less specific, using instead a more general expectation of positive results (see Bandura, 1977, 1986).

The present research synergized the three aspects by creating an HCM, but the poor convergent validity indicates that the HCM is incomplete. Much more variation in the general expectation (the first order) was attributed to unobserved variables than to the victim and situation (the second orders). The literature gives few clues, but enforcement of social standards (see Chekroun & Brauer, 2002; see O'Connell et al., 1999; see Salmivalli et al., 2011) and granting empathy (see Sainio et al., 2011) may be additional lower-ordered variables.

Of the two included dimensions, some literature shows that the victim-centric approach is stronger (see DeSmet et al., 2016), but the present results showed it to be weaker (see Table 33). The dichotomy may be due to the increased impact a flight attendant assault has on bystanders compared to ground-based incidents. Passenger-bystanders may feel the consequences of the disruption to everyone on board surpasses those incurred by the victim.

General Self-Efficacy. The *self-efficacy* outer model, designed by Schwarzer in 1999, was the only one designed exclusively for general use. It displayed the strongest

internal consistency and discriminant validity but had the poorest indicator reliability and convergent validity, although the latter two were not extreme. Even though the indicator reliability was low, it was at least stable, meaning there were no overly high or low-reliability indicators—all were comparable. Whether these characteristics are isolated to this research is unknown since an exhaustive evaluation of the model has not been published.

Taking Responsibility. In isolation, the *taking responsibility* outer model performed well regarding indicator reliability, internal consistency, and convergent validity. However, it performed poorly when correlated to other variables in the structural model (discriminant validity). Although the inference from such results is not definite, the relationship is unlikely to be due to measurement design. Instead, during a flight attendant assault, step three of the BIM is apparently intrinsically very similar to its following two steps. Such a relationship has not been seen in any other BIM research, so it presents a unique occurrence.

It would mean that passengers feel as if there is an expectation or responsibility that comes with knowing what to do or say to stop the assault. In other words, a level of responsibility comes with knowing what to do. The traditional BIM has those two steps occurring in series, but in the present research, they may be linked together and occur in parallel. There was a similar relationship with *deciding to act*.

According to discriminant validity results, the element of responsibility also shares much of its meaning with *deciding to act*—the final step of the BIM. Similar to the above reasoning, passengers may see deciding to act and taking personal responsibility as somewhat synonymous. In the traditional BIM, a bystander may feel responsible but yet

still decide not to act. In the flight attendant assault BIM, the feeling of personal responsibility and the decision to act would be inseparably linked—one could not occur without the other.

The Remaining Measurement Models. The remaining outer models and their constructs (*perceived social influence*, *recognizing the need for intervention*, *knowing what to do*, and *deciding to act*) performed well regarding reliability and validity and showed no particularly noteworthy characteristics. Each had been taken from parallel bystander research (see Albayrak-Aydemir & Gleibs, 2021; see Burn, 2009; see Pozzoli & Gini, 2010) and modified, but for the most part showed no degradation in usefulness due to doing so. Two constructs (*recognizing the need for intervention* and *deciding to act*) had slightly low internal consistency—falling just below the threshold (see Table 24). In its source literature, the internal consistency of *deciding to act* was at a similar level (α = .7) (Burn, 2009), but *recognizing the need for intervention* was considerably higher (α = .88 - .93) (Albayrak-Aydemir & Gleibs, 2021). The latter difference is likely attributed to the population and intervention scenario difference.

Structural Model Relationships

Four of the hypothesized relationships (H_1, H_2, H_5, and H_8) are not included in the following analysis because *taking responsibility* was removed from the model. The remaining structural relationships are discussed below.

Recognizing the Need for Intervention Influencing Knowing What to Do. The limited research into the BIM is consistent in finding that each step influences its following step (see Albayrak-Aydemir & Gleibs, 2021; see Anker & Feeley, 2011; see Christy & Voigt, 1994; see Nickerson et al., 2014; see Rabow et al., 1990). However,

bystander theory does not support step two influencing step four. There is also a gap in research related to why any step of the BIM would influence subsequent steps. It would be logical to assume that the influence of one step would not be so narrow as to cease after its following step. However, such an idea has only scantly been observed in literature (see Anker & Feeley, 2011) and is misaligned with BIM theory.

According to the results of this research, if a step influences past its following step, then the ultimate decision to act could be reached even if some BIM steps were skipped. However, BIM theory does not support this, which holds that failing to achieve any single step stops the BIM flow (Latané & Darley, 1968, 1970). Thus, the relationship between *recognizing the need for intervention* and *knowing what to do* may be a type I error. A more likely explanation for the present results, which aligns with BIM theory, is that, as discussed above, in a flight attendant assault, the step *taking responsibility* (step 3) is inherent in *knowing what to do* (step 4) and *deciding to act* (step 5). In other words, step three is not entirely missing from the modified model (see Figure 8) but incorporated into its following two steps. As a result, the positive relationship between step two and step four is due to the inherent inclusion of step three in step four.

An alternate explanation is that recognizing the urgency of an assault on an authority figure may include an inherent understanding of the appropriate response to reestablish that authority. In other words, when an authority figure is the victim, the relationship between steps two and four may exist whether or not step three occurs. The authority figure being victimized may make understanding what to do more readily apparent than when the victim is not one in authority. Although speculative, both

explanations would account for the significance of the relationship and the high correlation between indicators (high HTMT).

Knowing What to Do Influencing Deciding to Act. If bystanders do not know what actions to take, they will not decide to intervene (Latané & Darley, 1968, 1970). The relationship between the two steps is common (see Christy & Voigt, 1994; see Nickerson et al., 2014; see Rabow et al., 1990) but not universally found in all scenarios (see Albayrak-Aydemir & Gleibs, 2021). The difference appears to be dictated by either the bystander having first-hand awareness of the event or by the event being a single occurrence (as opposed to an enduring systemic crisis). The relationship was significant in this research for both males and females and aligns with the above distinction since the flight attendant assault would be a single event witnessed by the passenger. However, it may also mean that the group of passengers available to help is limited to only those who see or hear the event themselves.

Nonintervention Cost Influencing Recognizing the Need for Intervention. The *nonintervention cost* construct introduced in this research was novel. In traditional bystander literature, the victim exclusively feels the burden of nonintervention (see Fischer et al., 2011; see Liebst, 2019). In this research, the target of the nonintervention cost was not the victim but the bystander. That fact means there is also a gap in research regarding its influence on BIM steps. At the macro level, anecdotal evidence indicates that a high cost of nonintervention leads to more helping behavior (see Fischer et al., 2011; see Liebst, 2019). The results of this research support that finding but expand it to include the burden experienced by the bystander. In addition, the results of this research create more specificity as to which step of the BIM is the target of the influence.

Out-Group Social Bias Influencing Recognizing the Need for Intervention. Out-group social bias as a predictor of *recognizing the need for intervention* is not found in the literature. The relationship was specified after the removal of *taking responsibility*. Surprisingly, it was the only relationship in which males and females were completely homogeneous, both in the two outer models and the relationship between them. The results mean that the positive bias toward flight attendants is characterized equally for both males and females, influencing both groups to similarly interpret the state of need.

The relationship supports previous findings that deservedness and worthiness can influence a bystander to act (see Benson et al., 1976; see Burn, 2009; see Piliavin et al., 1969, 1975). If passengers have a positive bias toward flight attendants, they may view them as being more deserving and worthy of help, and this may manifest itself in a stronger belief that the event necessitates intervention. In other words, the flight attendant needs help because they deserve it. Researchers have yet to evaluate the influence of social bias on this step of the BIM, making the present findings unique.

Intervention Skills Influencing Knowing What to Do. The strongest relationship in the model, for both males and females, is intervention skills positively influencing a bystander to know how to respond. The strength aligns with prior findings that, in specific contexts, lacking such skills is one of the most significant causes of a bystander remaining passive (Burn, 2009). The results also confirm prior findings that the skills need not be specific to the event but can be of a general helping nature (see Huston et al., 1981).

Self-Efficacy Influencing Knowing What to Do. Bystander research utilizing the self-efficacy variable overwhelmingly specializes efficacy to the conflict scenario

(see DeSmet et al., 2016; see Pöyhönen et al., 2012; see Gini et al., 2008). This research, however, utilizes general self-efficacy—a belief in one's ability to achieve desired results in a broad context (Bandura, 1997). The results indicate that even though it lacks specificity, general self-efficacy can also influence a person's belief that they can handle a specific event—such as a flight attendant assault. Statistically, however, the present findings are only valid for males since the female sample size was inadequate to ensure the resultant path coefficient (β = .144) was not due to type I error.

Perceived Social Influence Influencing Deciding to Act. People's behavior can be influenced by those with whom the person has close personal ties (Caravita et al., 2009; Cillessen & Rose, 2005; Espelage et al., 2003; Gini, 2006, 2007; Goossens et al., 2006; Juvonen & Galvan, 2008; Lease et al., 2002; Newcomb et al., 1993; Nickerson & Mele-Taylor, 2014; Parkhurst & Hopmeyer, 1998; Salmivalli et al., 1996; Sandstrom & Cillessen, 2006; Slee, 1994). The pressure can even extend to intervention behavior, influencing a person to act even if they feel no personal responsibility to do so (Pozzoli & Gini, 2010, 2013; Rigby & Johnson, 2006). The present results support those previous findings for both males and females, with no significant difference between the two groups.

Expectation of Positive Outcome Influencing Deciding to Act. When a person intervenes, they have some expectation that doing so will have a prosocial result (Bandura, 1977, 1986). Two often-found aspects of the expectation are those centered on the victim (see Sainio et al., 2011) and the situation—maintaining a social standard (see Chekroun & Brauer, 2002; see O'Connell et al., 1999; see Salmivalli et al., 2011). A pro-victim attitude correlating with increased helping is very common (see Davis, 1994; see

DeSmet et al., 2016; see Eisenberg, 1991; see Eisenberg & Fabes, 1998; see Endresen & Olweus, 2001; see Gini et al., 2007, 2008; see Hoffman, 2001; see Philpot et al., 2020; see Pöyhönen et al., 2012; see Rudolph et al., 2004; see Weiner, 1980). The relationship is even more pronounced when the event is violent (Hart & Miethe, 2008). The same relationship is also found among bystanders who desire more to stop errant behavior (Dillon & Bushman, 2015; Pöyhönen et al., 2012). However, neither relationship appeared to be relevant in a flight attendant assault. Even when restricting the expectation dimensions to only those mentioned above (victim and situation), a correlation between such attitudes and the helping decision did not materialize for males or females. It was the only theorized relationship that failed to do so.

The reasons for the nonsignificant relationship between *expectation of positive outcome* and *deciding to act* are speculative but may be due to the unique characteristics of the airline cabin. Without the option to flee during a disturbance, passengers may be much more concerned with maintaining social tranquility than with ensuring the safety of a single individual or upholding social norms. If a disturbance continues, passengers may become more concerned with how they are personally affected. That concern may alter their perceived role. The ongoing disturbance may make them feel as if they are the one being victimized, even more so than the target of the assault. In their mind, the passenger may transition from being a bystander to a victim. The existence of such a shift has thus far been unresearched, possibly because in no other scenario is the bystander confined in the event.

Research Questions

The researcher used all results to answer the two research questions. The first is: Which steps of the bystander intervention model represent the relationships between existing theoretical factors and a passenger's willingness to intervene during an inflight assault on a flight attendant? After removing step three (*taking responsibility*) from the BIM, steps two (*recognizing the need for intervention*) and four (*knowing what to do*) showed significant ability to influence and predict their following step. However, the literature or bystander theory does not support step two influencing step four. Notwithstanding the discussion above, the BIM cannot, in its entirety, be applied to a flight attendant assault. Bystander theory includes all steps of the BIM, but this research did not support that. That fact makes its full applicability to a flight attendant assault suspect.

Still, the BIM appears to be at least partially helpful for understanding passenger-bystanders. However, the poor validity of *taking responsibility* means that the applicability of the BIM to an inflight assault cannot be fully assessed until the research community further explores the step of *taking responsibility*. The high HTMT between taking responsibility and other endogenous variables would typically lead to combining indicators or creating an HCM (see Hair et al., 2022). However, future researchers should be cautious since neither is supported by the literature, conceptually or statistically (see Aydemir & Gleibs, 2021; see Nickerson et al., 2014). Only with more evidence can researchers thoroughly assess whether practitioners can wholly apply the BIM as theorized to a flight attendant assault or if a new behavioral model is more appropriate.

The variety of results between males and females is noteworthy. Some constructs showed full invariance, others partial, and others showed high variance. Such a variety means practitioners and researchers alike cannot consider males and females to be either always homogeneous or always heterogeneous. Researchers should carefully evaluate each practical idea (construct), the influence of those ideas (path coefficient), and the impact (R^2) before applying them universally to either group. Additionally, of the four paths in which males and females could be compared, only one showed a significant difference (*nonintervention cost* to *recognizing the need for intervention*). The path difference means the personal negative impact of nonintervention influences males to apply a greater severity to a flight attendant assault than females do. The difference aligns with similar bystander research showing a higher tendency of males to intervene, particularly in severe or dangerous situations (see Austin, 1979; see Belansky & Boggiano, 1994; see Eagley & Crowley, 1986; see Liebst et al., 2019; see Shotland & Heinold, 1985). All other paths could either not be compared (due to group variance) or showed no difference between males and females. Overall, this illustrates how complex the gender variable is in bystander scenarios and research.

Research question two is: Which factors significantly predict a passenger's willingness to intervene during an inflight assault on a flight attendant? *Expectation of positive outcome* did not significantly predict *deciding to act*, although *perceived social influence* did. The distinction between the two exogenous variables provides evidence that social image amongst friends and family influences the helping impulse more than the desire to make things better for the victim or reduce such incidents—aligning with the self-centered bystander theory discussed in Chapter II.

All other exogenous variables significantly predicted their target constructs. However, most relationships are generally weak, being closer to zero than to a value of one. Of course, *weak* and *strong* are relative terms. Without a complete list of all influential variables, it is difficult to conclude the relative levels compared to those unspecified. *Intervention skills* had the most prominent influence for males and females, and *knowing what to do* had the highest amount of explained variance for both groups—over 50% for females. Even though readers cannot compare the two groups in this context, the fact that the relationship and explained variance ranked highest for both males and females indicates the importance of a bystander having some helping skills—even if not directly related to the incident. The results give some credibility to the efficacy of intervention training and the idea that a propensity to intervene can be a learned response.

Participant Recommendations

Few participants added comments to the open-ended question at the survey's close, but those who did presented some novel ideas. Some participants indicated that the fear of arrest by law enforcement mitigated their desire to intervene. Although participants did not reveal reasons for such fears, they appear to be driven by recent incidents in which bystanders or even victims were charged with assault or attempted murder for defensive actions. The fear introduces a construct completely absent in bystander literature. Fear of reprisal due to intervening is usually attributed to the attacker, but no research has included fear of retaliation from government authorities. It is a new but perhaps relevant phenomenon.

While evaluating *nonintervention cost* content validity, some SMEs identified other potential dimensions of the construct. Although too few SMEs identified dimensions enough to warrant inclusion, some are worthy of note. One SME identified a reduced perception of the airline and air travel as a possible added dimension to nonintervention cost. Corporate image or product degradation is not found in the literature, although it would only personally impact the bystander if they were a stakeholder. Two SMEs identified dimensions related to the long-term impacts of nonintervention, though both were unique. One was the lasting regret a bystander may feel if they do not intervene, and the other related to the lasting effect on children who would witness the assault. Even though the two dimensions are different, the idea of a bystander intervening to avoid a long-term effect is also not found in literature and may be worthy of consideration.

Conclusions

Theoretical Contributions

Only one path coefficient (*expectation of positive outcome* to *deciding to act*) was nonsignificant, making the other five outer models relevant passenger-bystander influencers during a flight attendant assault. However, except in one case (females *knowing what to do*), at least half of the explained variance of the endogenous constructs was due to unspecified variables. The low R^2 potentially means fewer influential variables were specified than were not specified, highlighting a remaining gap in passenger-bystander research.

Out-group social bias influencing *recognizing the need for intervention* presented a relationship not found in the literature. The relationship's novelty, significance, and

homogeneity (between males and females) create a new understanding of how preconceptions can influence a person's interpretation of an event. Positive bias begins to influence a bystander long before they decide whether or not to act. The finding creates a central point from which the boundaries of bias influence can be explored. The outer models' reliability and validity also make them useful for follow-on research—in particular the constructs *nonintervention cost* and *out-group social bias*.

Even though the dimensions of *nonintervention cost* were airline-specific, the construct introduced a novel perspective on the consequences of passivity by attributing them to the bystander and not the victim. Likewise, the construct *out-group social bias* added a positive dimension to a concept almost universally seen as unfavorable. A person's interpretation of an event can be affected by negative *and positive* perceptions of its actors. These new dimensions need not be unique to a flight attendant assault but can apply to other contexts.

The core BIM model (outer and inner) was estimated positively despite the discriminant validity problems of *taking responsibility*. Its relevance to a flight attendant assault remains debatable, however, since a valid measurement of the errant step three in this context has yet to be discovered. However, based on the significance of the remaining relationships, the model should be retained in passenger-bystander research.

Practical Contributions

The most influential variable in the model was *intervention skills*. Of particular note was that the skills dimensions in the construct (law enforcement, life-saving, first-aid, self-defense, and medical) were not directly related to two-person conflict resolution but were broader—intervention in an emergency. The detail seems to imply that having

any emergency-response training will boost the likelihood that a person will come to the aid of a flight attendant, or at least that they will be more likely to know what to do (BIM step four). While practitioners should not necessarily alter procedures based on these results alone, the implications are that when preselecting potential helpers, flight attendants should not simply look for indicators of combative type skills but instead for any emergency response training.

Recommendations

Stakeholders

Practitioners should continue to explore methods for eliciting more passenger intervention by increasing the accomplishment of each step in the intervention model. Flight attendants conveying a message of need (step two) may elicit intervention before an event becomes more violent. In addition, flight attendants should be familiarized with bystander theory, at least to the point of understanding the inhibitions to intervene and the process to overcome them. However, stakeholders should not isolate their attention to only flight attendants.

Passenger-bystanders are only one member of the assault triad. Even though this research was limited to bystander behavior, practitioners should not ignore the other two actors—the attacker and the flight attendant. Strategies for reducing the impact of inflight assaults should not just focus on eliciting bystander intervention but also on preventing and coping. The origin of an assault may begin long before boarding (Nelms, 1998; see DeCelles et al., 2019; see Genç & Dural, 2009; see McLinton et al., 2020; see Richards et al., 2016; see Vredenburgh et al., 2015), and the effects may remain long after deplaning, particularly for females (see Williams, 2000). These two phases should not be ignored.

Future Research

Much of the research community has ignored the social dynamics of flight attendants and the airline cabin, so those topics present significant opportunities for research. Specific to the model in this study, researchers should explore additional exogenous variables to create a more complete understanding of bystander motivations. The measurement of *taking responsibility* should also receive attention to either clarify its contribution or further justify its exclusion. The use of alternate measurement models may prove valuable in developing a behavioral model more suited to a flight attendant assault. Researchers should also not rule out the possibility that the BIM is an inaccurate representation of bystander behavior during an assault, and they should continue to explore new and perhaps more accurate behavioral models.

Broadly concerning flight attendant assaults, qualitative research is needed to understand the phenomenon from the perspectives of all those in the assault triad. The results of such research may allow for either refinement of the BIM or the generation of a unique model tailored to the specifics of airline travel. Research into the expectations the various stakeholders (flight attendants, passengers, managers, federal air marshals) have of each other concerning flight attendant assaults may reveal disconnects and opportunities to increase cooperation. Researchers have the option of adopting either quantitative or qualitative methods for such research, as both may provide unique and equally useful findings. Also, based on the open-ended question, the variable "fear of reprisal from law enforcement agencies" should be especially explored since it may represent a new and potentially significant negative influencer of bystander behavior.

Summary

The chapter reviewed the problems encountered during data collection and the challenges with determining the content validity of formative constructs. It presented an in-depth analysis of measurement models and structural relationships, and answers to the research questions were discussed. Theoretical and practical contributions were also explained. The chapter closed by making recommendations for stakeholders to reduce and mitigate the effects of assaults and for future researchers to continue to add to the body of knowledge.

References

Abbott, N., & Cameron, L. (2014). What makes a young assertive bystander? The effect of intergroup contact, empathy, cultural openness, and in-group bias on assertive bystander intervention intentions. *Journal of Social Issues, 70*(1), 167-182. https://doi.org/10.1111/josi.12053

Aboud, F. (1988). *Children and prejudice*. Blackwell.

Acuna, E., & Rodriguez, C. (2004, February 15-18). The treatment of missing values and its effect on classifier accuracy. In *Classification, Clustering, and Data Mining Applications: Proceedings of the Meeting of the International Federation of Classification Societies (IFCS)*. (pp. 639-647). Springer. https://doi.org/10.1007/978-3-642-17103-1_60

Akgeyik, T. (2011). Air rage: Violence toward cabin crew (A study on victimization of unruly passengers in Turkey). *Review of Business Research, 11*(3), 68-73.

Alam, S., Ayub, M. S., Arora, S., & Khan, M. A. (2023). An investigation of the imputation techniques for missing values in ordinal data enhancing clustering and classification analysis validity. *Decision Analytics Journal, 9*. https://doi.org/10.1016/j.dajour.2023.100341

Albayrak-Aydemir, N., & Gleibs, I. H. (2021). Measuring global bystander intervention and exploring its antecedents for helping refugees. *The British Journal of Psychology, 112*(2), 519–548. https://doi.org/10.1111/bjop.12474

Allen, H. (1968). *An exploratory study into bystander intervention and helping behavior in the subway* [Doctoral dissertation, New York University]. ProQuest Dissertations Publishing.

Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy, 15*(2), 214-221. https://doi.org/10.1016/j.sapharm.2018.03.066

Amazon Mechanical Turk (2019, April 18). Qualifications and worker task quality [Blog]. *Happenings at MTurk*. https://blog.mturk.com/qualifications-and-worker-task-quality-best-practices-886f1f4e03fc

Anderson, E. (2000). *Code of the street: Decency, violence, and the moral life of the inner city*. WW Norton & Company.

Andreou, E., & Metallidou, P. (2004). The relationship of academic and social cognition to behaviour in bullying situations among Greek primary school children. *Educational Psychology, 24*(1), 27-41. https://doi.org/10.1080/0144341032000146421

Anglin, L., Neves, P., Giesbrecht, N., & Kobus-Matthews, M. (2003). Alcohol-related air rage: From damage control to primary prevention. *The Journal of Primary Prevention, 23*(3), 283–297. https://doi.org/10.1023/A:1021341707993

Anker, A. E., & Feeley, T. H. (2011). Are nonparticipants in prosocial behavior merely innocent bystanders?. *Health Communication*, *26*(1), 13-24. https://doi.org/10.1080/10410236.2011.527618

Aratani, L. (2016, December 18). *Flight and fight: Attendants learn self-defense in the air*. The Washington Post. https://www.washingtonpost.com/local/trafficandcommuting/flight-and-fight-attendants-learn-self-defense- in-the-air/2016/12/18/b3ca3074-bc97-11e6-ac85-094a21c44abc_story.html

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology, 63*(4), 596–612. https://doi.org/10.1037/0022-3514.63.4.596

Ashton, N. L., & Severy, L. J. (1976). Arousal and costs in bystander intervention. *Personality and Social Psychology Bulletin, 2*(3). 268-272. https://doi.org/10.1177/014616727600200313

Associated Press. (2021, July 13). *Charlotte-bound passenger restrained with tape, cuffs after biting flight attendant, airline says*. WSOC TV. https://www.msn.com/en-us/travel/news/charlotte-bound-passenger-restrained-with-tape-cuffs-after-biting-flight-attendant-airline-says/ar-AAM6xGs

Atlas, R. S., & Pepler, D. J. (1998). Observations of bullying in the classroom. *The Journal of Educational Research (Washington, D.C.), 92*(2), 86–99. https://doi.org/10.1080/00220679809597580

Austin, W. (1979). Sex differences in bystander intervention in a theft. *Journal of Personality and Social Psychology*, *37*(11), 2110–2120. https://doi.org/10.1037/0022-3514.37.11.2110

Bailey, J. (2022, September 17). *What happens during a medical emergency during a flight?*. Simpleflying. https://simpleflying.com/medical-emergency-on-flight/

Ballard, T. J., Romito, P., Lauria, L., Vigiliano, V., Caldora, M., Mazzanti, C., & Verdecchia, A. (2006). Self perceived health and mental health among women flight attendants. *Occupational and Environmental Medicine, 63*(1), 33-38. https://doi.org/10.1136/oem.2004.018812

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215. https://doi.org/10.1037//0033-295X.84.2.191

Bandura, A. (1986). *Social foundations of thought and action*. Prentice-Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Macmillan.

Banyard, V. L. (2008). Measurement and correlates of prosocial bystander behavior: The case of interpersonal violence. *Violence and Victims*, *23*(1), 83–97. https://doi.org/10.1891/0886-6708.23.1.83

Banyard, V. L., Moynihan, M. M., Cares, A. C., & Warner, R. (2014). How do we know if it works? Measuring outcomes in bystander-focused abuse prevention on campuses. *Psychology of Violence*, *4*(1), 101-115. https://doi.org/10.1037/a0033470

Banyard, V. L., Moynihan, M. M., & Plante, E. G. (2007). Sexual violence prevention through bystander education: An experimental evaluation. *Journal of Community Psychology*, *35*(4), 463–481. https://doi.org/10.1002/jcop.20159

Baranishyn, M., Cudmore, B. A., & Fletcher, T. (2010). Customer service in the face of flight delays. *Journal of Vacation Marketing, 16*(3), 201-215. https://doi.org/10.1177/1356766710373681

Barclay, D. W., Thompson, R. & Higgins, C. (1995). The partial least squares (PLS) approach to causal modeling: Personal computer use as an illustration. *Technology Studies, 2*(2), 285-309.

Barron, P. E. (2002). Air rage: An emerging challenge for the airline industry. *Asia Pacific Journal of Transport, 4*(Summer), 39–44.

Batson, C. D. (1998). Altruism and prosocial behavior. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 2, pp. 282–316). McGraw-Hill/Oxford University Press.

Batson, C. D., Fultz, J., & Schoenrade, P. A. (1987). Distress and empathy: Two qualitatively distinct vicarious emotions with different motivational consequences. *Journal of Personality*, *55*(1), 19-39. https://doi.org/10.1111/j.1467-6494.1987.tb00426.x

Belansky, E. S., & Boggiano, A. K. (1994). Predicting helping behaviors: The role of gender and instrumental/expressive self-schemata. *Sex Roles*, *30*(9), 647-661. https://doi.org/10.1007/BF01544668

Bell, K. D. (2022). An evaluation into the causes of perpetual disruptive passenger behavior. *Journal of Transportation Security, 15*(1-2), 1–22. https://doi.org/10.1007/s12198-021-00243-5

Benson, P. L., Karabenick, S. A., & Lerner, R. M. (1976). Pretty pleases: The effects of physical attractiveness, race, and sex on receiving help. *Journal of Experimental Social Psychology*, *12*(5), 409-415. https://doi.org/10.1016/0022-1031(76)90073-1

Berkley, B. J., & Ala, M. (2001). Identifying and controlling threatening airline passengers. *Cornell Hospitality Quarterly, 42*(4), 6–23. https://doi.org/10.1177/0010880401424001

Berkowitz, L. (1972). Social norms, feelings and other factors affecting helping and altruism. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 63-108). Academic Press. https://doi.org/10.1016/S0065-2601(08)60025-8

Berkowitz, L., & Daniels, L. R. (1963). Responsibility and dependency. *The Journal of Abnormal and Social Psychology, 66*(5), 429–436. https://doi.org/10.1037/h0049250

Bethke, T. M., & DeJoy, D. M. (1993). An experimental study of factors influencing the acceptability of dating violence. *Journal of Interpersonal Violence, 8*(1), 36-51. https://doi.org/10.1177/088626093008001003

Bettencourt, B., Charlton, K., Dorr, N., & Hume, D. L. (2001). Status differences and in-group bias: A meta-analytic examination of the effects of status stability, status legitimacy, and group permeability. *Psychological Bulletin, 127*(4), 520-542. https://doi.org/10.1037/0033-2909.127.4.520

Bickman, L. (1972). Social influence and diffusion of responsibility in an emergency. *Journal of Experimental Social Psychology, 8*(5), 438–445. https://doi.org/10.1016/0022-1031(72)90069-8

Bihm, E., Gaudet, I., & Sale, O. (1979). Altruistic responses under conditions of anonymity. *The Journal of Social Psychology*, *109*(1), 25–30. https://doi.org/10.1080/00224545.1979.9933635

Black, D. (1993). *The social structure of right and wrong*. Academic Press.

Blanz, M., Mummendey, A., & Otten, S. (1995). Perceptions of relative group size and group status: Effects on intergroup discrimination in negative evaluations. *European Journal of Social Psychology*, *25*(2), 231-247. https://doi.org/10.1002/ejsp.2420250208

Blanz, M., Mummendey, A., & Otten, S. (1997). Normative evaluations and frequency expectations regarding positive versus negative outcome allocations between groups. *European Journal of Social Psychology*, *27*(2), 165-176. https://doi.org/10.1002/(SICI)1099-0992(199703)27:2<165::AID-EJSP812>3.0.CO;2-3

Boldry, J. G., & Kashy, D. A. (1999). Intergroup perception in naturally occurring groups of differential status: A social relations perspective. *Journal of Personality and Social Psychology*, *77*(6), 1200-1212. https://doi.org/10.1037/0022-3514.77.6.1200

Boomsma, A. (1982). Robustness of LISREL against small sample sizes in factor analysis models. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observations, causality, structure, prediction (Part 1)* (pp. 149-173). North Holland.

Bor, R. (2003). Trends in disruptive passenger behaviour on board UK registered aircraft: 1999–2003. *Travel Medicine and Infectious Disease, 1*(3), 153-157. https://doi.org/10.1016/j.tmaid.2003.09.006

Bor, R. (2007). Psychological factors in airline passenger and crew behaviour: A clinical overview. *Travel Medicine and Infectious Disease, 5*(4), 207–216. https://doi.org/10.1016/j.tmaid.2007.03.003

Bor, R., Russell, M., Parker, J., & Papadopoulos, L. (2001). Survey of world's airlines highlights various approaches to handling disruptive passengers. *ICAO Journal, 56*(2), 21-23.

Bor, R., & Van Gerwen, L. (2003). *Psychological perspectives on fear of flying*. Ashgate.

Braithwaite, R. (2001). *Managing aggression*. Routledge. https://doi.org/10.4324/9780203193914

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin, 86*(2), 307–324. https://doi.org/10.1037/0033-2909.86.2.307

Brewer M. B. (2001). Ingroup identification and intergroup conflict: When does ingroup love become outgroup hate? In R. Ashmore, L. Jussim, & D. Wilder (Eds.), *Social identity, intergroup conflict, and conflict reduction* (pp. 17-41). Oxford University Press. https://doi.org/10.1093/oso/9780195137422.003.0002

Brewer, M. B., & Campbell, D. T. (1976). *Ethnocentrism and intergroup attitudes: East African evidence.* Sage.

Brewin, C. R. (1984). Perceived controllability of life-events and willingness to prescribe psychotropic drugs. *British Journal of Social Psychology, 23*(3), 285–287. https://doi.org/10.1111/j.2044-8309.1984.tb00642.x

Brewster, M., & Tucker, J. M. (2016). Understanding bystander behavior: The influence of and interaction between bystander characteristics and situational factors. *Victims & Offenders, 11*(3), 455-481. https://doi.org/10.1080/15564886.2015.1009593

Bricker, J. B. (2005). Development and evaluation of the air travel stress scale. *Journal of Counseling Psychology, 52*(4), 615-628. https://doi.org/10.1037/0022-0167.52.4.615

Brown, G., Widing, R. E., & Coulter, R. L. (1991). Customer evaluation of retail salespeople utilizing the SOCO scale: A replication, extension, and application. *Journal of the Academy of Marketing Science, 19*(4), 347–351. https://doi.org/10.1007/BF02726510

Brigham, J. C. (1986). *Social psychology.* Little Brown.

Bryan, J. H., & London, P. (1970). Altruistic behavior by children. *Psychological Bulletin, 73*(3), 200–211. https://doi.org/10.1037/h0028744

Bryan, J. H., & Test, M. A. (1967). Models and helping: Naturalistic studies in aiding behavior. *Journal of Personality and Social Psychology, 6*(4p1), 400–407. https://doi.org/10.1037/h0024826

Budd, L. C. S. (2011). On being aeromobile: Airline passengers and the affective experiences of flight. *Journal of Transport Geography, 19*(5), 1010–1016. https://doi.org/10.1016/j.jtrangeo.2010.06.017

Bukowski, W., Newcomb, A., & Hartup, W. (Eds.). (1996). *The company they keep: Friendship in childhood and adolescence*. Cambridge University Press.

Burn, S. M. (2009). A situational model of sexual assault prevention through bystander intervention. *Sex roles*, *60*(11), 779-792. https://doi.org/10.1007/s11199-008-9581-5

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed.). Routledge. https://doi.org/10.4324/9781315757421

Cameron, L., Rutland, A., Brown, R., & Douch, R. (2006). changing children's intergroup attitudes toward refugees: Testing different models of extended contact. *Child Development*, *77*(5), 1208–1219. https://doi.org/10.1111/j.1467-8624.2006.00929.x

Caprara, G. V., & Steca, P. (2005). Self-efficacy beliefs as determinants of prosocial behavior conducive to life satisfaction across ages. *Journal of Social and Clinical Psychology*, *24*(2). https://doi.org/10.1521/jscp.24.2.191.62271

Caprara, G. V., Steca, P., Cervone, D., & Artistico, D. (2003). The contribution of self-efficacy beliefs to dispositional shyness: On social-cognitive systems and the development of personality dispositions. *Journal of Personality, 71*(6), 943-970. https://doi.org/10.1111/1467-6494.7106003

Caravita, S. C., Di Blasio, P., & Salmivalli, C. (2009). Unique and interactive effects of empathy and social status on involvement in bullying. *Social Development*, *18*(1), 140-163. https://doi.org/10.1111/j.1467-9507.2008.00465.x

Carlson, M. (2008). I'd rather go along and be considered a man: Masculinity and bystander intervention. *The Journal of Men's Studies 16*(1), 3–17. https://doi.org/10.3149/jms.1601.3

Carney, J. V., Hazler, R. J., Oh, I., Hibel, L. C., & Granger, D. A. (2010). The relations between bullying exposures in middle childhood, anxiety, and adrenocortical activity. *Journal of School Violence, 9*(2), 194–211. https://doi.org/10.1080/15388220903479602

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's Mturk, social media, and face-to-face behavioral testing. *Computers in Human Behavior, 29*(6), 2156-2160. https://doi.org/10.1016/j.chb.2013.05.009

CBSLA Staff. (2021, May 25). *Southwest Airlines attendant assaulted by unruly passenger on flight from Sacramento to San Diego*. CBS Los Angeles. https://losangeles.cbslocal.com/2021/05/25/southwest-airlines-attendant-assaulted-by-unruly-passenger-on-flight-from-sacramento-to-san-diego/

CBSNewYork. (2016, January 8). *New York to Chicago flight diverted when woman allegedly attacks flight attendant*. CBSNewYork. https://newyork.cbslocal.com/2016/01/08/laguardia-chicago-flight-diverted/

Chang, L. (1994). A psychometric evaluation of four-point and six-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement, 18*(3), 205-215. https://doi.org/10.1177/014662169401800302

Chang, L. (2003). Variable effects of children's aggression, social withdrawal, and prosocial leadership as functions of teacher beliefs and behaviors. *Child Development, 74(2)*, 535-548. https://doi.org/10.1111/1467-8624.7402014

Chekroun, P. & Brauer, M. (2002). The bystander effect and social control behavior: The effect of the presence of others on people's reactions to norm violations. *European Journal of Social Psychology*, *32*(6), 853–867. https://doi.org/10.1002/ejsp.126

Chen, X., Yu, H., & Yu, F. (2015). What is the optimal number of response alternatives for rating scales? From an information processing perspective. *Journal of Marketing Analytics, 3*(2), 69-78. https://doi.org/10.1057/jma.2015.4

Christy, C. A., & Voigt, H. (1994). Bystander responses to public episodes of child abuse. *Journal of Applied Social Psychology*, *24*(9), 824-847. https://doi.org/10.1111/j.1559-1816.1994.tb00614.x

Chyung, S.Y., Roberts, K., Swanson, I., & Hankinson, A. (2017). Evidence-based survey design: The use of a midpoint on the Likert scale. *Performance Improvement, 56*(10), 15-23. http://dx.doi.org/10.1002/pfi.21727

Cialdini, R. B., Brown, S. L., Lewis, B. P., Luce, C., & Neuberg, S. L. (1997). Reinterpreting the empathy–altruism relationship: When one into one equals oneness. *Journal of Personality and Social Psychology*, *73*(3), 481-494. https://doi.org/10.1037/0022-3514.73.3.481

Cillessen, A. H., & Mayeux, L. (2004). From censure to reinforcement: Developmental changes in the association between aggression and social status. *Child Development*, *75*(1), 147-163. https://doi.org/10.1111/j.1467-8624.2004.00660.x

Cillessen, A. H., & Rose, A. J. (2005). Understanding popularity in the peer system. *Current Directions in Psychological Science*, *14*(2), 102-105. https://doi.org/10.1111/j.0963-7214.2005.00343.x

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309–319. https://doi.org/10.1037/1040-3590.7.3.309

Clark, R. D., & Word, L. E. (1974). Where is the apathetic bystander? Situational characteristics of the emergency. *Journal of Personality and Social Psychology*, *29*(3), 279-287. https://doi.org/10.1037/h0036000

Clark, R. D., & Word, L. E. (1972). Why don't bystanders help? Because of ambiguity? *Journal of Personality and Social Psychology*, *24*(3), 392–400. https://doi.org/10.1037/h0033717

Clasen, D. R., & Brown, B. B. (1985). The multidimensionality of peer pressure in adolescence. *Journal of youth and adolescence*, *14*(6), 451-468. https://doi.org/10.1007/BF02139520

Coie, J. D., & Koeppl, G. K. (1990). Adapting intervention to the problems of aggressive and disruptive children. In S. R. Asher & J. D. Coie (Eds.), *Peer rejection in childhood* (pp. 309-337). Cambridge University Press.

Colman, A. M., Norris, C. E., & Preston, C. C. (1997). Comparing rating scales of different lengths: Equivalence of scores from 5-point and 7-point scales. *Psychological Reports*, *80*(2), 355–362. https://doi.org/10.2466/pr0.1997.80.2.355

Cook, C. C. H. (1997). Alcohol and aviation. *Addiction*, *92*(5), 539–555. https://doi.org/10.1111/j.1360-0443.1997.tb02912.x

Cowie, H. (2000). Bystanding or standing by: Gender issues in coping with bullying in English schools. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *26*(1), 85-97. https://doi.org/10.1002/(SICI)1098-2337(2000)26:1<85::AID-AB7>3.0.CO;2-5

Craig, W. M. & Pepler, D. J. (1995). Peer processes in bullying and victimization: An observational study. *Exceptionality Education Canada*, *5*, 81-95. https://doi.org/10.1007/BF02128600

Cramer, R. E., McMaster, M. R., Bartell, P. A., & Dragna, M. (1988). Subject competence and minimization of the bystander effect. *Journal of Applied Social Psychology*, *18*(13), 1133-1148. https://doi.org/10.1111/j.1559-1816.1988.tb01198.x

Cunningham, C. E., Cunningham, L. J., Martorelli, V., Tran, A., Young, J., & Zacharias, R. (1998). The effects of primary division, student-mediated conflict resolution programs on playground aggression. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *39*(5), 653-662. https://doi.org/10.1017/S0021963098002558

Dahlberg, A. (2016). *Air rage: The underestimated safety risk*. Routledge. https://doi.org/10.4324/9781315263083

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

Davis, M. H. (1994). *Empathy: A social psychological approach*. Brown & Benchmark.

DeCarlo, E. (1998, 11 May). Murder, she read. *New York Magazine, 31*(18), 38-41.

DeCelles, K. A., DeVoe, S. E., Rafaeli, A., & Agasi, S. (2019). Helping to reduce fights before flights: How environmental stressors in organizations shape customer emotions and customer–employee interactions. *Personnel Psychology, 72*(1), 49-80. https://doi.org/10.1111/peps.12292

DeCelles, K. A., & Norton, M. I. (2016). Physical and situational inequality on airplanes predicts air rage. *Proceedings of the National Academy of Sciences*, *113*(20), 5588-5591. https://doi.org/10.1073/pnas.1521727113

DeSmet, A., Bastiaensens, S., Van Cleemput, K., Poels, K., Vandebosch, H., Cardon, G., & De Bourdeaudhuij, I. (2016). Deciding whether to look after them, to like it, or leave it: A multidimensional analysis of predictors of positive and negative bystander behavior in cyberbullying among adolescents. *Computers in Human Behavior*, *57*, 398–415. https://doi.org/10.1016/j.chb.2015.12.051

Dickson, S. (2021). *Administrator's letter to airport leaders on the zero tolerance policy*. Federal Aviation Administration. https://www.faa.gov/data_research/passengers_cargo/unruly_passengers/toolkit/media/Letter_to_airports_FINAL_signed.pdf

Dijkstra, T. K., & Henseler, J. (2015). Consistent partial least squares path modeling. *MIS Quarterly*, *39*(2), 297–316. https://doi.org/10.25300/MISQ/2015/39.2.02

Dillon, K. P., & Bushman, B. J. (2015). Unresponsive or un-noticed?: Cyberbystander intervention in an experimental cyberbullying context. *Computers in Human Behavior*, *45*, 144–150. https://doi.org/10.1016/j.chb.2014.12.009

Dion, K. L. (1973). Cohesiveness as a determinant of ingroup-outgroup bias. *Journal of Personality and Social Psychology*, *28*(2), 163-171. https://doi.org/10.1037/h0035725

Doise, W., Csepeli, G., Dann, H. D., Gouge, C., Larsen, K., & Ostell, A. (1972). An experimental investigation into the formation of intergroup representations. *European Journal of Social Psychology*, *2*(2), 202-204. https://doi.org/10.1002/ejsp.2420020208

Dovidio, J. F., Piliavin, J. A., Gaertner, S. L., Schroeder, D. A., & Clark, R. D., III (1991). The arousal/cost-reward model and the process of intervention: A review of the evidence. In M. S. Clark (Ed.), *Prosocial behavior* (pp. 86–118). Sage.

Dovidio, J. F., Piliavin, J. A., Schroeder, D. A., & Penner, L. A. (2017). *The social psychology of prosocial behavior*. Psychology Press. https://doi.org/10.4324/9781315085241

Duckitt, J. H. (1992). *The social psychology* of *prejudice*. Praeger.

Eagly, A. H., & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin*, *100*(3), 283–308. https://doi.org/10.1037/0033-2909.100.3.283

Eisenberg, N. (1991). Meta-analytic contributions to the literature on prosocial behavior. *Personality and Social Psychology Bulletin*, *17*(3), 273-282. https://doi.org/10.1177/0146167291173007

Eisenberg, N., & Eggum, N. D. (2009). Empathic responding: Sympathy and personal distress. In J. Decety & W. J. Ickes (Eds.), *The social neuroscience of empathy* (pp. 71–84). MIT Press. https://doi.org/10.7551/mitpress/9780262012973.003.0007

Eisenberg, N. & Fabes, R. A. (1998). Prosocial development. In W. Damon (Editor in Chief), & N. Eisenberg (Vol. Ed.), *Social, emotional, and personality. Handbook of child psychology* (Vol. 3, pp. 701–778). Wiley.

Eisenberg, N., Guthrie, I. K., Cumberland, A., Murphy, B. C., Shepard, S. A., Zhou, Q., & Carlo, G. (2002). Prosocial development in early adulthood: A longitudinal study. *Journal of Personality and Social Psychology*, *82*(6), 993–1006. https://doi.org/10.1037/0022-3514.82.6.993

Ellemers, N., Wilke, H., & Van Knippenberg, A. (1993). Effects of the legitimacy of low group or individual status on individual and collective status-enhancement

strategies. *Journal of Personality and Social Psychology*, *64*(5), 766-778. https://doi.org/10.1037/0022-3514.64.5.766

Embry-Riddle Aeronautical University. (2023a). *Participant recruitment and consent.* https://erau.edu/research/irb/informed-consent

Embry-Riddle Aeronautical University. (2023b). *Definitions and categories of research.* https://erau.edu/research/irb/definitions-and-categories-of-research

Emswiller, T., Deaux, K., & Willits, J. E. (1971). Similarity, sex, and requests for small favors. *Journal of Applied Social Psychology*, *1*(3), 284-291. https://doi.org/10.1111/j.1559-1816.1971.tb00367.x

Endresen, I. M. & Olweus, D. (2001). Self-reported empathy in Norwegian adolescents: Sex differences, age trends, and relationship to bullying. In A. C. Bohart & D. J. Stipek (Eds.), *Constructive & destructive behavior: Implications for family, school, & society* (pp. 147–165). American Psychological Association. https://doi.org/10.1037/10433-007

Espelage, D. L., Holt, M. K., & Henkel, R. R. (2003). Examination of peer–group contextual effects on aggression during early adolescence. *Child Development*, *74*(1), 205-220. https://doi.org/10.1111/1467-8624.00531

European Union Aviation Safety Agency. (n.d.). *Not on my flight: Zero tolerance against unruly passengers* [Video]. YouTube. https://www.youtube.com/watch?v=5YF0P7SLPtk

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobstrusive measure of racial attitudes: A bona fide pipeline? J*ournal of Personality and Social Psychology, 69*(6), 1013–1027. https://doi.org/10.1037//0022-3514.69.6.1013

Federal Aviation Administration. (2021a). *2021 Unruly passenger data.* https://www.faa.gov/data_research/passengers_cargo/unruly_passengers/2021_archive

Federal Aviation Administration. (2021b). *2021 Airport digital signage.* https://www.faa.gov/data_research/passengers_cargo/unruly_passengers/toolkit/digital_signage

Federal Aviation Administration. (n.d.). *2021 Dangerous behavior doesn't fly.* https://www.faa.gov/unruly

Feinman, S. (1978). When does sex affect altruistic response?. *Psychological Reports, 43*(3). https://doi.org/10.2466/pr0.1978.43.3f.1218

Fischer, P., & Greitemeyer, T. (2013). The Positive bystander effect: Passive bystanders increase helping in situations with high expected negative consequences for the

helper. *The Journal of Social Psychology*, *153*(1), 1–5. https://doi.org/10.1080/00224545.2012.697931

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

Fitz-Gibbon, J. (2021, July 20). *Reports of unruly U.S. airline passengers up 500% this year*. New York Post. Retrieved October 12, 2021 from https://nypost.com/2021/07/20/reports-of-unruly-us-airline-passengers-up-500-this-year/

Fleiss, J. (1981). *Statistical methods for rates and proportions* (2nd ed.). John Wiley.

Gaertner, S. L., & Dovidio, J. F. (1977). The subtlety of White racism, arousal, and helping behavior. *Journal of Personality and Social Psychology*, *35*(10), 691–707. https://doi.org/10.1037/0022-3514.35.10.691

Gaertner, S. L., Dovidio, J. F., & Johnson, G. (1982). Race of victim, nonresponsive bystanders, and helping behavior. *The Journal of Social Psychology*, *117*(1), 69–77. https://doi.org/10.1080/00224545.1982.9713409

Gale, S., Mordukhovich, I., Newlan, S., & McNeely, E. (2018). The impact of verbal abuse, sexual harassment and sexual assault at work: An investigation into the health of flight attendants in the Harvard Flight Attendant Health Study. *The Lancet*, (August 13). https://doi.org/10.2139/ssrn.3234856

Gallo, M. M. (2015). *"No one helped": Kitty Genovese, New York City, and the myth of urban apathy*. Cornell University Press. https://doi.org/10.7591/9780801455902

Garcia, S. M., Weaver, K., Moskowitz, G. B., & Darley, J. M. (2002). Crowded minds: The implicit bystander effect. *Journal of Personality and Social Psychology*, *83*(4), 843-853. https://doi.org/10.1037/0022-3514.83.4.843

Gardham, K., & Brown, R. (2001). Two forms of intergroup discrimination with positive and negative outcomes: Explaining the positive-negative asymmetry effect. *British Journal of Social Psychology*, *40*(1), 23-34. https://doi.org/10.1348/014466601164678

Genç, B., & Dural, U. (2009). Flight anxiety, coping and management. *British Travel Health Association Journal, 14*, 26-32.

George, D., Carroll, P., Kersnick, R., & Calderon, K. (1998). Gender-related patterns of helping among friends. *Psychology of Women Quarterly*, *22*(4), 685-704. https://doi.org/10.1111/j.1471-6402.1998.tb00185.x

Gini, G. (2006). Bullying as a social process: The role of group membership in students' perception of inter-group aggression at school. *Journal of School Psychology*, *44*(1), 51-65. https://doi.org/10.1016/j.jsp.2005.12.002

Gini, G. (2007). Who is blameworthy?: Social identity and inter-group bullying. *School Psychology International*, *28*(1), 77-89. https://doi.org/10.1177/0143034307075682

Gini, G., Albiero, P., Benelli, B., & Altoe, G. (2007). Does empathy predict adolescents' bullying and defending behavior?. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *33*(5), 467-476. https://doi.org/10.1002/ab.20204

Gini, G., Albiero, P., Benelli, B., & Altoe, G. (2008). Determinants of adolescents' active defending and passive bystanding behavior in bullying. *Journal of Adolescence*, *31*(1), 93-105. https://doi.org/10.1016/j.adolescence.2007.05.002

Girasek, D. C., & Olsen, C. H. (2009). Airline passengers' alcohol use and its safety implications. *Journal of Travel Medicine, 16*(5), 311–316. https://doi.org/10.1111/j.1708-8305.2009.00339.x

Goffman, E. (1959). *The presentation of self in everyday life.* Doubleday.

Goossens, F. A., Olthof, T., & Dekker, P. H. (2006). New participant role scales: Comparison between various criteria for assigning roles and indications for their validity. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *32*(4), 343-357. https://doi.org/10.1002/ab.20133

Gottlieb, J., & Carver, C. S. (1980). Anticipation of future interaction and the bystander effect. *Journal of Experimental Social Psychology, 16*(3), 253–260. https://doi.org/10.1016/0022-1031(80)90068-2

Graziano, W. G., & Habashi, M. M. (2010). Motivational processes underlying both prejudice and helping. *Personality and Social Psychology Review*, *14*(3), 313-331. https://doi.org/10.1177/1088868310361239

Graziano, W. G. & Habashi, M. M. (2015). Searching for the prosocial personality. In D. A. Schroeder & W. G. Graziano (Eds.), *The Oxford Handbook of Prosocial Behavior* (pp. 231–255). Oxford University Press.

Greitemeyer, T., Fischer, P., Kastenmüller, A., & Frey, D. (2006). Civil courage and helping behavior: Differences and similarities. *European Psychologist*, *11*(2), 90-98. https://doi.org/10.1027/1016-9040.11.2.90

Greitemeyer, T., & Rudolph, U. (2003). Help giving and aggression from an attributional perspective: Why and when we help or retaliate. *Journal of Applied Social Psychology*, *33*(5), 1069-1087. https://doi.org/10.1111/j.1559-1816.2003.tb01939.x

Gross, R. D. (1994). *Key studies in psychology* (2nd ed.). St. Martin's Press.

Gruder, C. L., & Cook, T. D. (1971). Sex, dependency, and helping. *Journal of Personality and Social Psychology*, *19*(3), 290-294. https://doi.org/10.1037/h0031456

Grzyb, T. (2016). Why can't we just ask? The influence of research methods on results. The case of the "bystander effect." *Polish Psychological Bulletin*, *47*(2), 233–235. https://doi.org/10.1515/ppb-2016-0027

Gusfield, J., Kotarba, J., & Rasmussen, P. (1981). The public society of intimates: Friends, wives, lovers, and others in the drinking-driving drama. *Research in the Interweave of Social Roles*, *2*, 237-257.

Guy, R. F., & Norvell, M. (1977). The neutral point on a Likert scale. *The Journal of Psychology:Interdisciplinary and Applied*, *95*(2), 199–204. https://doi.org/10.1080/00223980.1977.9915880

Hagendoorn, L. (1995). Intergroup biases in multiple group systems: The perception of ethnic hierarchies. *European Review of Social Psychology*, *6*(1), 199-228. https://doi.org/10.1080/14792779443000058

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Prentice Hall.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). Sage. https://doi.org/10.1007/978-3-030-80519-7

Hair, J. F., Sarstedt, M., Hopkins, L., & Kuppelwieser, V. G. (2014). Partial least squares structural equation modeling (PLS-SEM) an emerging tool in business research. *European Business Review*, *26*(2), 106-121. https://doi.org/10.1108/EBR-10-2013-0128

Hair, J. F., Sarstedt, M., Ringle, C. M., & Gudergan, S. P. (2018). *Advanced issues in partial least squares structural equation modeling*. Sage. https://doi.org/10.1007/978-3-319-05542-8_15-1

Hall, J. A. (1990). *Nonverbal sex differences: Accuracy of communication and expressive style*. Johns Hopkins University Press.

Harada, J. (1985). Bystander intervention The effect of ambiguity of the helping situation and the interpersonal relationship between bystanders. *Japanese Psychological Research, 27*(4), 177-184. https://doi.org/10.4992/psycholres1954.27.177

Harari, H., Harari, O., & White, R. V. (1985). The reaction to rape by American male bystanders. *The Journal of Social Psychology, 125*(5), 653–658. https://doi.org/10.1080/00224545.1985.9712039

Hardy, S. A., & Carlo, G. (2005). Religiosity and prosocial behaviours in adolescence: The mediating role of prosocial values. *Journal of Moral Education, 34*(2), 231–249. https://doi.org/10.1080/03057240500127210

Harms, P. D., & DeSimone, J. A. (2015). Caution! MTurk workers ahead—fines doubled. *Industrial and Organizational Psychology, 8*(2), 183-190. https://doi.org/10.1017/iop.2015.23

Harris, V. A., & Robinson, C. E. (1973). Bystander intervention: Group size and victim status. *Bulletin of the Psychonomic Society, 2*(1), 8-10. https://doi.org/10.3758/BF03327696

Hart, T. C., & Miethe, T. D. (2008). Exploring bystander presence and intervention in nonfatal violent victimization: When does helping really help? *Violence and Victims, 23*(5), 637–651. https://doi.org/10.1891/0886-6708.23.5.637

Hastings, P. D., Zahn-Waxler, C., Robinson, J., Usher, B., & Bridges, D. (2000). The development of concern for others in children with behavior problems. *Developmental Psychology, 36*(5), 531–546. https://doi.org/10.1037/0012-1649.36.5.531

Hauser, D. J., & Schwarz, N. (2016). Attentive turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods, 48*(1), 400-407. https://doi.org/10.3758/s13428-015-0578-z

Hazler, R. J. (1996). *Breaking the cycle of violence: Interventions for bullying and victimization*. Taylor & Francis.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Henseler, J., Ringle, C. M., & Sarstedt, M. (2016). Testing measurement invariance of composites using partial least squares. *International Marketing Review, 33*, 405-431. https://doi.org/10.1108/IMR-09-2014-0304

Henseler, J., Ringle, C. M., & Sinkovics, R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing, 20*, 277-320. https://doi.org/10.1108/S1474-7979(2009)0000020014

Hibbitts, B. (2006, April 12). *Flight 93 cockpit voice recorder transcript [US DC]*. Jurist. https://www.jurist.org/news/2006/04/flight-93-cockpit-voice-recorder/

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, *53*(1), 575–604. https://doi.org/10.1146/annurev.psych.53.100901.135109

Hoefnagels, C., & Zwikker, M. (2001). The bystander dilemma and child abuse: Extending the Latane and Darley model to domestic violence. *Journal of Applied Social Psychology*, *31*(6), 1158-1183. https://doi.org/10.1111/j.1559-1816.2001.tb02668.x

Hoffman, M. L. (1977). Sex differences in empathy and related behaviors. *Psychological Bulletin*, *84*(4), 712–722. https://doi.org/10.1037/0033-2909.84.4.712

Hoffman, M. L. (2001). Toward a comprehensive empathy-based theory of prosocial moral development. In A. C. Bohart & D. J. Stipek (Eds.), *Constructive & destructive behavior: Implications for family, school, & society* (pp. 61–86). American Psychological Association. https://doi.org/10.1037/10433-003

Hortensius, R., & de Gelder, B. (2018). From empathy to apathy: The bystander effect revisited. *Current Directions in Psychological Science 27*(4), 249-256. https://doi.org/10.1177/0963721417749653

Horowitz, I. A. (1971). The effect of group norms on bystander intervention. *The Journal of Social Psychology*, *83*(2), 265-273. https://doi.org/10.1080/00224545.1971.9922471

Howard, W., & Crano, W. D. (1974). Effects of sex, conversation, location, and size of observer group on bystander intervention in a high risk situation. *Sociometry*, 491-507. https://doi.org/10.2307/2786423

Hu, H-H., Hu, H-Y., & King, B. (2017). Impacts of misbehaving air passengers on frontline employees: Role stress and emotional labor. *International Journal of Contemporary Hospitality Management*, *29*(7), 1793-1813. https://doi.org/10.1108/IJCHM-09-2015-0457

Hunter, J. A. (2006). A correlational study of how airline customer service and consumer perception of airline customer service affect the air rage phenomenon. *Journal of Air Transportation*, *11*(3), 78-109. https://ntrs.nasa.gov/citations/20070022531

Hunter, S. C., & Borg, M. G. (2006). The influence of emotional reaction on help seeking by victims of school bullying. *Educational Psychology*, *26*(6), 813-826. https://doi.org/10.1080/01443410600941946

Hunter, S. C., Boyle, J. M., & Warden, D. (2004). Help seeking amongst child and adolescent victims of peer-aggression and bullying: The influence of school-stage,

gender, victimization, appraisal, and emotion. *British Journal of Educational Psychology*, *74*(3), 375-390. https://doi.org/10.1348/0007099041552378

Hurley, D., & Allen, B. P. (1974). The effect of the number of people present in a nonemergency situation. *The Journal of Social Psychology, 92*(1), 27-29. https://doi.org/10.1080/00224545.1974.9923068

Huston, T. & Korte, C. (1976). The responsive bystander: Why he helps. In T. Lickona (Ed.), *Moral development and behavior* (pp. 269-283). Holt, Rinehart, and Winston.

Huston, T. L., Ruggiero, M., Conner, R., & Geis, G. (1981). Bystander intervention into crime: A study based on naturally-occurring episodes. *Social Psychology Quarterly, 44*(1), 14-23. https://doi.org/10.2307/3033858

International Civil Organization Association. (2019). *Manual on the legal aspects of unruly and disruptive passengers* (1st ed.) (Report No. 10117).

Janos, A. (2020, September 8). *How United fight 93 passengers fought back on 9/11*. History Stories. https://www.history.com/news/united-flight-93-september-11-passengers-revolt-crash

Jenkins, L. N., & Nickerson, A. B. (2017). Bullying participant roles and gender as predictors of bystander intervention. *Aggressive Behavior, 43*(3), 281-290. https://doi.org/10.1002/ab.21688

Johns, R. (2005). One size doesn't fit all: Selecting response scales for attitude items. *Journal of Elections, Public Opinion and Parties, 15*(2), 237-264, https://doi.org/10.1080/13689880500178849

Jost, J. T. (2001). Outgroup favoritism and the theory of system justification: A paradigm for investigating the effects of socioeconomic success on stereotype content. In G. Moskowitz (Ed.), *Cognitive social psychology: The Princeton symposium on the legacy and future of social cognition* (pp. 89-102). Psychology Press. https://doi.org/10.4324/9781410605887

Jouriles, E. N., Kleinsasser, A., Rosenfield, D., & McDonald, R. (2016). Measuring bystander behavior to prevent sexual violence: Moving beyond self reports. *Psychology of Violence*, *6*(1), 73-81. https://doi.org/10.1037/a0038230

Juvonen, J. & Galvan, A. (2008). Peer influence in involuntary social groups: Lessons from research on bullying. In M. J. Prinstein & K. A. Dodge (Eds.), *Understanding peer influence in children and adolescents* (pp. 225–244). Guilford.

Kalafat, J., Elias, M., & Gara, M. A. (1993). The relationship of bystander intervention variables to adolescent responses to suicidal peers. *The Journal of Primary Prevention, 13*(4), 231–244. https://doi.org/10.1007/BF01324560

Karakashian, L. M., Walter, M. I., Christopher, A. N., & Lucas, T. (2006). Fear of negative evaluation affects helping behavior: The bystander effect revisited. *North American Journal of Psychology*, *8*(1), 13-32. https://psycnet.apa.org/record/2006-04346-003

Katz, P. A. (1976). The acquisition of racial attitudes in children. In P. A. Katz (Ed.), *Towards the elimination of racism* (pp. 125–154). Pergamon. https://doi.org/10.1016/B978-0-08-018316-9.50012-6

Kazdin, A. E., & Bryan, J. H. (1971). Competence and volunteering. *Journal of Experimental Social Psychology*, *7*(1), 87-97. https://doi.org/10.1016/0022-1031(71)90058-8

Keefe, K. (1994). Perceptions of normative social pressure and attitudes toward alcohol use: Changes during adolescence. *Journal of Studies on Alcohol*, *55*(1), 46-54. https://doi.org/10.15288/jsa.1994.55.46

Kleinsasser, A., Jouriles, E. N., McDonald, R., & Rosenfield, D. (2015). An online bystander intervention program for the prevention of sexual violence. *Psychology of Violence, 5*(3), 227-235. https://doi.org/10.1037/a0037393

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). Guilford Press.

Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. *Information Systems Journal (Oxford, England)*, *28*(1), 227–261. https://doi.org/10.1111/isj.12131

Korte, C. (1971). Effects of individual responsibility and group communication on help-giving in an emergency. *Human Relations*, *24*(2), 149–159. https://doi.org/10.1177/001872677102400204

Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. P*sychological Bulletin, 140*(4), 1073–1137. https://doi.org/10.1037/a0035618

Krajicek, D. (2011, March 13). *The killing of Kitty Genovese: 47 years later, still holds sway over New Yorkers*. Daily News. Retrieved September 2, 2022 from https://www.nydailynews.com/news/crime/killing-kitty-genovese-47-years-holds-sway-new-yorkers-article-1.123912

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213–236. https://doi.org/10.1002/acp.2350050305

Krosnick, J. A., & Fabrigar, L. R. (1997). Designing rating scales for effective measurement in surveys. In L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C.

Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality* (pp. 141-164). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118490013

Kulas, J., Stachowski, A., & Haynes, B. (2008). Middle response functioning in Likert-responses to personality items. *Journal of Business & Psychology, 22*(3), 251-259. https://doi.org/10.1007/s10869-008-9064-2

Kunesh, A. (2022, April 17). How to keep your points and miles from expiring. *The points guy*. https://thepointsguy.com/guide/keep-points-and-miles-from-expiring/

Labhardt, D., Holdsworth, E., Brown, S., & Howat, D. (2017). You see but you do not observe: A review of bystander intervention and sexual assault on university campuses. *Aggression and Violent Behavior, 35*, 13-25. http://dx.doi.org/10.1016/j.avb.2017.05.005

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Laner, M. R., Benin, M. H., & Ventrone, N. A. (2001). Bystander attitudes toward victims of violence: Who's worth helping? *Deviant Behavior*, *22*(1), 23–42. https://doi.org/10.1080/016396201750065793

Latané, B. (1981). The psychology of social impact. *American Psychologist*, *36*(4), 343-356. https://doi.org/10.1037/0003-066X.36.4.343

Latané, B., & Dabbs, J. M. (1975). Sex, group size and helping in three cities. *Sociometry*, *38*(2), 180-194. https://doi.org/10.2307/2786599

Latané, 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. (1969). Bystander "apathy". *American Scientist*, *57*(2), 244-268.

Latané, B. & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?*. Appleton-Century-Crofts.

Latané, B. & Nida, S. (1981). Ten years of research on group size and helping. *Psychological Bulletin*, *89*(2), 308–324. https://doi.org/10.1037/0033-2909.89.2.308

Latané, B., & Rodin, J. (1969). A lady in distress: Inhibiting effects of friends and strangers on bystander intervention. *Journal of Experimental Social Psychology*, *5*(2), 189-202. https://doi.org/10.1016/0022-1031(69)90046-8

Lease, A. M., Kennedy, C. A., & Axelrod, J. L. (2002). Children's social constructions of popularity. *Social Development, 11*(1), 87-109. https://doi.org/10.1111/1467-9507.00188

Leone, R. M., Schipani-McLaughlin, A. M., Haikalis, M., & Parrott, D. J. (2020). The "white knight" effect: Benevolent sexism accounts for bystander intervention in party situations among high status men. *Psychology of Men & Masculinity, 21*(4), 704-709. https://doi.org/10.1037/men0000314

Leung, S.-O. (2011). A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point Likert scales. *Journal of Social Service Research, 37*(4), 412–421. https://doi.org/10.1080/01488376.2011.580697

Levine, M., Cassidy, C., Brazier, G., & Reicher, S. (2002). Self-categorization and bystander non-intervention: Two experimental studies. *Journal of Applied Social Psychology*, *32*(7), 1452–1463. https://doi.org/10.1111/j.1559-1816.2002.tb01446.x

Levine, M., & Crowther, S. (2008). The responsive bystander: How social group membership and group size can encourage as well as inhibit bystander intervention. *Journal of Personality and Social Psychology*, *95*(6), 1429–1439. https://doi.org/10.1037/a0012634

Levine, M., Lowe, R., Best, R., & Heim, D. (2012). 'We police it ourselves': Group processes in the escalation and regulation of violence in the night-time economy. *European Journal of Social Psychology*, *42*(7), 924–932. https://doi.org/10.1002/ejsp.1905

Levine, M., & Manning, R. (2013). Social identity, group processes, and helping in emergencies. *European Review of Social Psychology, 24*(1), 225–251. https://doi.org/10.1080/10463283.2014.892318

Levine, M., Philpot, R., & Kovalenko, A. G. (2020). Rethinking the bystander effect in violence reduction training programs. *Social Issues and Policy Review, 14*(1), 273-296. https://doi.org/10.1111/sipr.12063

Levine, M., Prosser, A., Evans, D., & Reicher, S. (2005). Identity and emergency intervention: How social group membership and inclusiveness of group boundaries shape helping behavior. *Personality and Social Psychology Bulletin*, *31*(4), 443-453. https://doi.org/10.1177/0146167204271651

Levine, M., Taylor, P. J., & Best, R. (2011). Third parties, violence, and conflict resolution: The role of group size and collective action in the microregulation of violence. *Psychological Science*, *22*(3), 406–412. https://doi.org/10.1177/0956797611398495

Levine, R. V., Martinez, T. S., Brase, G., & Sorenson, K. (1994). Helping in 36 U.S. cities. *Journal of Personality and Social Psychology, 67*(1), 69–82. https://doi.org/10.1037/0022-3514.67.1.69

Levy, P., Lundgren, D., Ansel, M., Fell, D., Fink, B., & McGrath, J. E. (1972). Bystander effect in a demand-without-threat situation. *Journal of Personality and Social Psychology, 24*(2), 166-171. https://doi.org/10.1037/h0033380

Liebst, L. S., Philpot, R., Bernasco, W., Dausel, K. L., Ejbye-Ernst, P., Nicolaisen, M. H., & Lindegaard, M. R. (2019). Social relations and presence of others predict bystander intervention: Evidence from violent incidents captured on CCTV. *Aggressive Behavior, 45*(6), 598-609. https://doi.org/10.1002/ab.21853

Loewenstein, G., & Small, D. A. (2007). The Scarecrow and the Tin Man: The vicissitudes of human sympathy and caring. *Review of General Psychology, 11*(2), 112-126. https://doi.org/10.1037/1089-2680.11.2.112

Lofland, J. (1969). *Deviance and identity.* Prentice Hall.

Lowe, R. D., Levine, M., Best, R. M., & Heim, D. (2012). Bystander reaction to women fighting: Developing a theory of intervention. *Journal of Interpersonal Violence, 27*(9), 1802–1826. https://doi.org/10.1177/0886260511430393

Luckenbill, D. F. (1977). Criminal homicide as a situated transaction. *Social Problems, 25*(2), 176-186. https://doi.org/10.1525/sp.1977.25.2.03a00050

Luckey, S. (2000). Air Rage in our increasingly violent world, airline pilots need training for encounters with enraged passengers. *Air Line Pilot, 69*(8), 18-21.

Ma, X. (2002). Bullying in middle school: Individual and school characteristics of victims and offenders. *School effectiveness and school improvement, 13*(1), 63-89. https://doi.org/10.1076/sesi.13.1.63.3438

Macháčková, H., Dedkova, L., Sevcikova, A., & Cerna, A. (2013). Bystanders' support of cyberbullied schoolmates. *Journal of Community & Applied Social Psychology, 23*(1), 25–36. https://doi.org/10.1002/casp.2135

Markey, P. M. (2000). Bystander intervention in computer-mediated communication. *Computers in Human Behavior, 16*(2), 183–188. https://doi.org/10.1016/S0747-5632(99)00056-4

Marsh, H. W., & Bailey, M. (1991). Confirmatory factor analyses of multitrait-multimethod data: A comparison of alternative models. *Applied Psychological Measurement, 15*(1), 47-70. https://doi.org/10.1177/014662169101500106

Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study 1: Reliability and validity. *Educational and Psychological Measurement, 31*(3), 657-674. https://doi.org/10.1177/001316447103100307

McIntosh, I. B., Swanson, V., Power, K. G., Raeside, F., & Dempster, C. (1998). Anxiety and health problems related to air travel. *Journal of Travel Medicine, 5*(4), 198-204. https://doi.org/10.1111/j.1708- 8305.1998.tb00507.x

McKelvie, S. J. (1978). Graphic rating scales: How many categories? *British Journal of Psychology, 69*(2), 185-202. https://doi.org/10.1111/j.2044-8295.1978.tb01647.x

McLinton, S. S., Drury, D., Masocha, S., Savelsberg, H., Martin, L., & Lushington, K. (2020). "Air rage": A systematic review of research on disruptive airline passenger behaviour 1985-2020. *Journal of Airline and Airport Management, 10*(1), 31–49. https://doi.org/10.3926/jairm.156

Meldrum, R. C. (2016). Low self-control takes flight: The association between indicators of low self-control and imprudent airline passenger behavior. *The Social Science Journal, 53*(4), 444-454. https://doi.org/10.1016/j.soscij.2016.08.005

Menesini, E., Codecasa, E., Benelli, B., & Cowie, H. (2003). Enhancing children's responsibility to take action against bullying: Evaluation of a befriending intervention in Italian middle schools. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *29*(1), 1-14. https://doi.org/10.1002/ab.80012

Menon, K., & Dubé, L. (2004). Service provider responses to anxious and angry customers: Different challenges, different payoffs. *Journal of Retailing, 80*(3), 229-237. https://doi.org/10.1016/j.jretai.2003.11.002

Menon, K., & Dubé, L. (2007). The effect of emotional provider support on angry versus anxious consumers. *International Journal of Research in Marketing, 24*(3), 268-275. https://doi.org/10.1016/j.ijresmar.2007.04.001

Merrens, M. R. (1973). Nonemergency helping behavior in various sized communities. *The Journal of Social Psychology*, *90*(2), 327-328. https://doi.org/10.1080/00224545.1973.9712577

Meyer, J. P., & Mulherin, A. (1980). From attribution to helping: An analysis of the mediating effects of affect and expectancy. *Journal of Personality and Social Psychology*, *39*(2), 201-210. https://doi.org/10.1037/0022-3514.39.2.201

Michelini, R. L., Wilson, J. P., & Messé, L. A. (1975). The influence of psychological needs on helping behavior. *The Journal of Psychology*, *91*(2), 253-258. https://doi.org/10.1080/00223980.1975.9923950

Midlarsky, E. (1971). Aiding under stress: The effects of competence, dependency, visibility, and fatalism. *Journal of Personality*, *39*(1), 132-149. https://doi.org/10.1111/j.1467-6494.1971.tb00993.x

Midlarsky, E., & Midlarsky, M. (1973). Some determinants of aiding under experimentally-induced stress. *Journal of Personality*, *41*(2), 305-327. https://doi.org/10.1111/j.1467-6494.1973.tb00096.x

Miethe, T. D., & Deibert, G. R. (2006). *Fight time: The normative rules and routines of interpersonal violence*. Waveland Press Inc.

Miethe, T. D., Regoeczi, W. C., & Drass, K. A. (2004). *Rethinking homicide: Exploring the structure and process underlying deadly situations*. Cambridge University Press. https://doi.org/10.1017/CBO9780511499487

Milgram, S. (1970). The experience of living in cities. *Science, 167*(3924), 1461–1468. https://doi.org/10.1126/science.167.3924.1461

Monks, C., Ortega Ruiz, R., & Torrado Val, E. (2002). Unjustified aggression in preschool. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *28*(6), 458-476. https://doi.org/10.1002/ab.10032

Montagu, M. A. (1949). The origin and nature of social life and the biological basis of cooperation. *The Journal of Social Psychology*, *29*(2), 267-283. https://doi.org/10.1080/00224545.1949.9921791

Morgan, C. J. (1978). Bystander intervention: Experimental test of a formal model. *Journal of Personality and Social Psychology*, *36*(1), 43–55. https://doi.org/10.1037/0022-3514.36.1.43

Morgan, M., & Nickson, D. (2001). Uncivil aviation: A review of the air rage phenomenon. *The International Journal of Tourism Research*, *3*(6), 443–457. https://doi.org/10.1002/jtr.327

Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status: An integration. *European Journal of Social Psychology*, *22*(2), 103-122. https://doi.org/10.1002/ejsp.2420220202

Mummendey, A., & Otten, S. (1998). Positive–Negative asymmetry in social discrimination. *European Review of Social Psychology*, *9*(1), 107-143. https://doi.org/10.1177/0146167200262007

Mummendey, A., Simon, B., Dietze, C., Grünert, M., Haeger, G., Kessler, S., Lettgen, S., & Schäferhoff, S. (1992). Categorization is not enough: Intergroup discrimination in negative outcome allocation. *Journal of Experimental Social Psychology*, *28*(2), 125-144. https://doi.org/10.1016/0022-1031(92)90035-I

Nadler, J. T., Weston, R., & Voyles, E. C. (2015). Stuck in the middle: The use and interpretation of mid-points in items on questionnaires. *The Journal of General Psychology*, *142*(2), 71-89. https://doi.org/10.1080/00221309.2014.994590

Nelms, D. W., (1998). As more passengers fill more aircraft, the airlines are finding the skies are decidedly less friendly. *Air Transport World (35)*3, 34-35, 38-39, 41-43.

Nelson, J. K., Dunn, K. M., & Paradies, Y. (2011). Bystander anti-racism: A review of the literature. *Analyses of Social Issues and Public Policy, 11*(1), 263-284. https://doi.org/10.1111/j.1530-2415.2011.01274.x

Nesdale, D. & Scarlett, M. (2004) Effects of group and contextual factors on pre-adolescent children's attitudes to school bullying. *International Journal of Behavioral Development, 28*(5), 428–434. https://doi.org/10.1080/01650250444000144

Newcomb, A. F., Bukowski, W. M., & Pattee, L. (1993). Children's peer relations: A meta-analytic review of popular, rejected, neglected, controversial, and average sociometric status. *Psychological Bulletin*, *113*(1), 99-128. https://doi.org/10.1037/0033-2909.113.1.99

Newman, O. (1973). *Defensible space; Crime prevention through urban design*. Macmillan Publishing.

Nickerson, A. B., Aloe, A. M., Livingston, J. A., & Feeley, T. H. (2014). Measurement of the bystander intervention model for bullying and sexual harassment. *Journal of Adolescence*, *37*(4), 391-400. https://doi.org/10.1016/j.adolescence.2014.03.003

Nickerson, A. B., Mele, D., & Princiotta, D. (2008). Attachment and empathy as predictors of roles as defenders or outsiders in bullying interactions. *Journal of School Psychology*, *46*(6), 687-703. https://doi.org/10.1016/j.jsp.2008.06.002

Nickerson, A. B., & Mele-Taylor, D. (2014). Empathetic responsiveness, group norms, and prosocial affiliations in bullying roles. *School Psychology Quarterly*, *29*(1), 99-109. https://doi.org/10.1037/spq0000052

Nishina, A., & Juvonen, J. (2005). Daily reports of witnessing and experiencing peer harassment in middle school. *Child Development, 76*(2), 435–450. https://doi.org/10.1111/j.1467-8624.2005.00855.x

Nunnally, J. C. (1967). *Psychometric theory* (2nd ed.). McGraw-Hill.

Oakes, M., & Bor, R. (2010). The psychology of fear of flying (part I): A critical evaluation of current perspectives on the nature, prevalence and etiology of fear of flying. *Travel Medicine and Infectious Disease, 8*(6), 327–338. https://doi.org/10.1016/j.tmaid.2010.10.001

Obermaier, M., Fawzi, N., & Koch, T. (2014). Bystanding or standing by? How the number of bystanders affects the intention to intervene in cyberbullying. *New Media & Society*, *18*(8), 1491-1507. http://dx.doi.org/10.1177/1461444814563519

O'Connell, P., Pepler, D., & Craig, W. (1999). Peer involvement in bullying: Insights and challenges for intervention. *Journal of Adolescence*, *22*(4), 437-452. https://doi.org/10.1006/jado.1999.0238

Oh, I. & Hazler, R. J. (2009). Contributions of personal and situational factors to bystanders' reactions to school bullying. *School Psychology International*, *30*(3), 291–310. https://doi.org/10.1177/0143034309106499

Ojala, K., & Nesdale, D. (2004). Bullying and social identity: The effects of group norms and distinctiveness threat on attitudes towards bullying. *British Journal of Developmental Psychology*, *22*(1), 19–35. https://doi.org/10.1348/026151004772901096

Olweus, D. (1993). *Bullying at school: What we know and what we can do*. Blackwell.

Orlando, C. M. (2020). Peer intervention with suicidal disclosures on social media: Does the bystander effect play a role? [Doctoral dissertation, University of South Carolina]. Scholar Commons. https://scholarcommons.sc.edu/etd/6060

Otten, S., Mummendey, A., & Blanz, M. (1996). Intergroup discrimination in positive and negative outcome allocations: Impact of stimulus valence, relative group status, and relative group size. *Personality and Social Psychology Bulletin, 22*(6), 568-581. https://doi.org/10.1177/0146167296226003

Pantin, H. M., & Carver, C. S. (1982). Induced competence and the bystander effect. *Journal of Applied Social Psychology*, *12*(2), 100-111. https://doi.org/10.1111/j.1559-1816.1982.tb00852.x

Parkhurst, J. T., & Hopmeyer, A. (1998). Sociometric popularity and peer-perceived popularity: Two distinct dimensions of peer status. *The Journal of Early Adolescence*, *18*(2), 125-144. https://doi.org/10.1177/0272431698018002001

Pawlowski, A. (2013, January 8). *How flight attendants deal with unruly passengers*. NBCNews. https://www.nbcnews.com/business/travel/how-flight-attendants-deal-unruly-passengers-flna1B7885636

Phillips, S., & Cooney, M. (2005). Aiding peace, abetting violence: Third parties and the management of conflict. *American Sociological Review, 70*(2), 334–354. https://doi.org/10.1177/000312240507000207

Philpot, R., Liebst, L. S., Levine, M., Bernasco, W., & Lindegaard, M. R. (2020). Would I be helped? Cross-National CCTV footage shows that intervention is the norm in public conflicts. *The American Psychologist*, *75*(1), 66-75. https://doi.org/10.1037/amp0000469

Philpot, R., Liebst, L. S., Møller, K. K., Lindegaard, M. R., & Levine, M. (2019). Capturing violence in the night-time economy: A review of established and

emerging methodologies. *Aggression and Violent Behavior, 46,* 56–65. https://doi.org/10.1016/j.avb.2019.02.004

Pierson, K., Power, Y., Marcus, A., & Dahlberg, A. (2007) Airline passenger misconduct: Management implications for physicians. *Aviation Space and Environmental Medicine, 78*(4), 361–367.

Piliavin, J. A., Dovidio, J. F., Gaertner, S. L., & Clark, R. D. (1981). *Emergency intervention*. Academic Press.

Piliavin, I. M., Piliavin, J. A., & Rodin, J. (1975). *Journal of Personality and Social Psychology*, *32*(3), 429–438. https://doi.org/10.1037/h0077092

Piliavin, I. M., Rodin, J., & Piliavin, J. A. (1969). Good Samaritanism: An underground phenomenon? *Journal of Personality and Social Psychology*, *13*(4), 289–299. https://doi.org/10.1037/h0028433

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, *75*(3), 811-832. https://doi.org/10.1037/0022-3514.75.3.811

Planty, M. (2002). *Third-party involvement in violent crime, 1993-99*. U.S. Department of Justice, Bureau of Justice Statistics. https://www.ojp.gov/ncjrs/virtual-library/abstracts/third-party-involvement-violent-crime-1993-99

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, *30*(4), 459–467. https://doi.org/10.1002/nur.20199

Pontell, H. N., Salinger, L. R., & Geis, G. (1983). Assaults in the air: Concerning attacks against flight attendants. *Deviant Behavior*, *4*(3-4), 297-311. https://doi.org/10.1080/01639625.1983.9967619

Potts, J., & Lynch, B. (2010). Stepping forward or shrinking back: Psychological factors in bystanders' decisions to act (or not). *Resuscitation*, *81*(2), S47–S47. https://doi.org/10.1016/j.resuscitation.2010.09.199

Pöyhönen, V., Juvonen, J., & Salmivalli, C. (2010). What does it take to stand up for the victim of bullying? The interplay between personal and social factors. *Merrill-Palmer Quarterly*, *56*(2), 143–163. https://doi.org/10.1353/mpq.0.0046

Pöyhönen, Juvonen, J., & Salmivalli, C. (2012). Standing up for the victim, siding with the bully or standing by? Bystander responses in bullying situations. *Social Development (Oxford, England)*, *21*(4), 722–741. https://doi.org/10.1111/j.1467-9507.2012.00662.x

Pöyhönen, V., & Salmivalli, C. (2008). New directions in research and practice addressing bullying: Focus on defending behavior. In D. Pepler & W. Craig

(Eds.), *Understanding and addressing bullying: An international perspective* (pp. 26-39). AuthorHouse.

Pozzoli, T., & Gini, G. (2010). Active defending and passive bystanding behavior in bullying: The role of personal characteristics and perceived peer pressure. *Journal of Abnormal Child Psychology*, *38*(6), 815-827. https://doi.org/10.1007/s10802-010-9399-9

Pozzoli, T., & Gini, G. (2013). Why Do Bystanders of Bullying Help or Not? A Multidimensional Model. *The Journal of Early Adolescence*, *33*(3), 315–340. https://doi.org/10.1177/0272431612440172

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica, 104*(2000), 1-15. https://doi.org/10.1016/S0001-6918(99)00050-5

Prinstein, M. J., & Dodge, K. A. (Eds.). (2008). *Understanding peer influence in children and adolescents*. Guilford Press.

Prohibition on Interference with Crew Members, 14 C.F.R. § 91.11. (2023). https://www.ecfr.gov/current/title-14/chapter-I/subchapter-F/part-91/subpart-A/section-91.11

Rabbie, J. M., & Horwitz, M. (1969). Arousal of ingroup-outgroup bias by a chance win or loss. *Journal of Personality and Social Psychology*, *13*(3), 269-277. https://doi.org/10.1037/h0028284

Rabow, J., Newcomb, M. D., Monto, M. A., & Hernandez, A. C. (1990). Altruism in drunk driving situations: Personal and situational factors in intervention. *Social Psychology Quarterly*, 199-213. https://doi.org/10.2307/2786959

Reichl, A. J. (1997). Ingroup favouritism and outgroup favouritism in low status minimal groups: Differential responses to status-related and status-unrelated measures. *European Journal of Social Psychology*, *27*(6), 617-633. https://doi.org/10.1002/(SICI)1099-0992(199711/12)27:6<617::AID-EJSP829>3.0.CO;2-T

Reinartz, W., Haenlein, M., & Henseler, J. (2009). An empirical comparison of the efficacy of covariance-based and variance-based SEM. *International Journal of Research in Marketing (26)*4, 332-344. https://doi.org/10.1016/j.ijresmar.2009.08.001

Rennie, D. (2021, October 28). *How James Bulger was killed by Jon Venables and Robert Thompson*. Allthatsinteresting. Retrieved September 1, 2022 from https://allthatsinteresting.com/james-bulger-case

Rhoden, S., Ralston, R., & Ineson, E. M. (2008). Cabin crew training to control disruptive airline passenger behavior: A cause for tourism concern? *Tourism Management, 29*(3), 538–547. https://doi.org/10.1016/j.tourman.2007.06.002

Rice, S., Winter, S. R., Doherty, S., & Milner, M. (2017). Advantages and disadvantages of using internet-based survey methods in aviation-related research. *Journal of Aviation Technology and Engineering, 7*(1), 58-65. http://dx.doi.org/10.7771/2159-6670.1160

Richards, P., Cleland, J., & Zuckerman, J. (2016). Psychological factors relating to physical health issues: How physical factors in aviation and travel affect psychological functioning. In T. Hubbard, & R. Bor (Eds.), *Aviation Mental Health* (pp. 27-38). Routledge. https://doi.org/10.4324/9781315568560

Rigby, K., & Johnson, B. (2006). Expressed readiness of Australian schoolchildren to act as bystanders in support of children who are being bullied. *Educational Psychology*, *26*(3), 425-440. https://doi.org/10.1080/01443410500342047

Rigby, K., & Slee, P. T. (1991). Bullying among Australian school children: Reported behavior and attitudes toward victims. *The Journal of Social Psychology*, *131*(5), 615-627. https://doi.org/10.1080/00224545.1991.9924646

Rigby, K., & Slee, P. T. (1993). Dimensions of interpersonal relation among Australian children and implications for psychological well-being. *The Journal of Social Psychology*, *133*(1), 33-42. https://doi.org/10.1080/00224545.1993.9712116

Ross, A. S. (1971). Effect of increased responsibility on bystander intervention: The presence of children. *Journal of Personality and Social Psychology*, *19*(3), 306–310. https://doi.org/10.1037/h0031459

Rudolph, U., Roesch, S., Greitemeyer, T., & Weiner, B. (2004). A meta-analytic review of help giving and aggression from an attributional perspective: Contributions to a general theory of motivation. *Cognition and Emotion*, *18*(6), 815-848. https://doi.org/10.1080/02699930341000248

Ruhl, C. (2021, April 20). *Kitty Genovese*. SimplyPsychology. Retrieved September 2, 2022 from https://www.simplypsychology.org/Kitty-Genovese.html

Rushton, J. P., & Campbell, A. C. (1977). Modeling, vicarious reinforcement and extraversion on blood donating in adults: Immediate and long-term effects. *European Journal of Social Psychology*, *7*(3), 297-306. https://doi.org/10.1002/ejsp.2420070304

Rutkowski, G. K., Gruder, C. L., & Romer, D. (1983). Group cohesiveness, social norms, and bystander intervention. *Journal of Personality and Social Psychology, 44*(3), 545–552. https://doi.org/10.1037/0022-3514.44.3.545

Rutland, A. (1999). The development of national prejudice, in-group favouritism and self-stereotypes in British children. *British Journal of Social Psychology, 38*(1), 55–70. https://doi.org/10.1348/014466699164031

Rutland, A., Cameron, L., Bennett, L., & Ferrell, J. (2005). Interracial contact and racial constancy: A multi-site study of racial intergroup bias in 3–5 year old Anglo-British children. *Journal of Applied Developmental Psychology, 26*(6), 699–713. https://doi.org/10.1016/j.appdev.2005.08.005

Sainio, M., Veenstra, R., Huitsing, G., & Salmivalli, C. (2011). Victims and their defenders: A dyadic approach. *International Journal of Behavioral Development*, *35*(2), 144-151. https://doi.org/10.1177/0165025410378068

Salinger, L. M., Jesilow, P., Pontell, H. N., & Geis, G. (1985). Assaults against airline flight attendants: A victimization study. *Transportation Journal, 25*(1), 66-71. http://www.jstor.org/stable/20712841

Salmivalli, C. (2010). Bullying and the peer group: A review. *Aggression and Violent Behavior, 15*(2), 112–120. https://doi.org/10.1016/j.avb.2009.08.007

Salmivalli, C., Lagerspetz, K., Björkqvist, K., Österman, K., & Kaukiainen, A. (1996). Bullying as a group process: Participant roles and their relations to social status within the group. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *22*(1), 1-15. https://doi.org/10.1002/(SICI)1098-2337(1996)22:1<1::AID-AB1>3.0.CO;2-T

Salmivalli, C., & Voeten, M. (2004). Connections between attitudes, group norms, and behaviour in bullying situations. *International Journal of Behavioral Development, 28*(3), 246-258. https://doi.org/10.1080/01650250344000488

Salmivalli, C., Voeten, M., & Poskiparta, E. (2011). Bystanders matter: Associations between reinforcing, defending, and the frequency of bullying behavior in classrooms. *Journal of Clinical Child and Adolescent Psychology, 40*(5), 668–676. https://doi.org/10.1080/15374416.2011.597090

Sandstrom, M. J., & Cillessen, A. H. (2006). Likeable versus popular: Distinct implications for adolescent adjustment. *International Journal of Behavioral Development, 30*(4), 305-314. https://doi.org/10.1177/0165025406072789

Sarstedt, M., Hair, J. F., Cheah, J. H., Becker, J. M., & Ringle, C. M. (2019). How to specify, estimate, and validate higher-order constructs in PLS-SEM. *Australian Marketing Journal, 27*(3), 197-211. https://doi.org/10.1016/j.ausmj.2019.05.003

Sarstedt, M., Henseler, J., & Ringle, C. M. (2011). Multi-group analysis in partial least squares (PLS) path modeling: Alternative methods and empirical results. In M. Sarstedt, M. Schwaiger, & C. R. Taylor (Eds.), *Advances in international*

marketing (Vol. 22, pp. 195–218). Emerald Group Publishing Limited. https://doi.org/10.1108/S1474-7979(2011)0000022012

Schwartz, S. H., & David, A. B. (1976). Responsibility and helping in an emergency: Effects of blame, ability and denial of responsibility. *Sociometry*, *39*(4), 406–415. https://doi.org/10.2307/3033505

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

Schwarzer, R. (1999, November 3). *General perceived self efficacy in 14 cultures.* http://userpage.fu-berlin.de/~gesund/publicat/ehps_cd/health/world14.htm

Schwarzer, R., Dunkel-Schetter, C., Weiner B., & Woo, G. (1992). Expectancies as mediators of the relationships between recipient characteristics and social support intentions. In R. Schwarzer (Ed.), *Self-efficacy: Thought control of action* (pp. 65-90). Taylor & Francis Group. https://doi.org/10.4324/9781315800820

Sedikides, C., Olsen, N., & Reis, H. T. (1993). Relationships as natural categories. *Journal of Personality and Social Psychology*, *64*(1), 71–82. https://doi.org/10.1037/0022-3514.64.1.71

Sheehan, K. B. (2018). Crowdsourcing research: Data collection with Amazon's Mechanical Turk. *Communication Monographs*, *85*(1), 140-156. https://doi.org/10.1080/03637751.2017.1342043

Shmueli, G., Sarstedt, M., Hair, J. F., Jun-Hwa Cheah, Ting, H., Vaithilingam, S., & Ringle, C. M. (2019). Predictive model assessment in PLS-SEM: Guidelines for using PLSpredict. *European Journal of Marketing*, *53*(11), 2322-2347. https://doi.org/10.1108/EJM-02-2019-0189

Shotland, R. L., & Heinold, W. D. (1985). Bystander response to arterial bleeding: Helping skills, the decision-making process, and differentiating the helping response. *Journal of Personality and Social Psychology*, *49*(2), 347-356. https://doi.org/10.1037/0022-3514.49.2.347

Shotland, R. L., & Huston, T. L. (1979). Emergencies: What are they and do they influence bystanders to intervene?. *Journal of Personality and Social Psychology*, *37*(10), 1822-1834. https://doi.org/10.1037/0022-3514.37.10.1822

Shotland, R. L., & Straw, M. K. (1976). Bystander response to an assault: When a man attacks a woman. *Journal of Personality and Social Psychology*, *34*(5), 990–999. https://doi.org/10.1037/0022-3514.34.5.990

Slee, P. T. (1994). Situational and interpersonal correlates of anxiety associated with peer victimization. *Child Psychiatry and Human Development*, *25*(2), 97-107. https://doi.org/10.1007/BF02253289

Small, D. A., & Loewenstein, G. (2003). Helping a victim or helping the victim: Altruism and identifiability. *Journal of Risk and Uncertainty*, *26*(1), 5-16. http://www.jstor.org/stable/41761095

Small, J., & Harris, C. (2018, February 5-8). Air rage: Who is responsible for unruly behaviour? [Conference presentation]. CAUTHE 2018: *Get Smart: Paradoxes and Possibilities in Tourism, Hospitality and Events Education and Research* (pp. 710-714). New Castle, New South Wales, Australia.

Smart, R. G., & Mann, R. E. (2003). Causes and consequences of air rage in Canada. *Canadian Journal of Public Health*, *94*(4), 251-253. https://doi.org/10.1007/BF03403544

Smith, E. R., & Henry, S. (1996). An in-group becomes part of the self: Response time evidence. *Personality & Social Psychology Bulletin, 22*(6), 635–642. https://doi.org/10.1177/0146167296226008

Soper, D. (2022). A-priori sample size calculator for structural equation models. *Free statistics calculators*. https://www.danielsoper.com/statcalc/calculator.aspx?id=89

Spells, A., Joseph, E., & Sanchez, R. (2021, June 12). *Delta passengers and crew subdue off-duty flight attendant on Atlanta-bound flight*. CNN. https://www.cnn.com/2021/06/12/us/delta-unruly-passenger-flight-diverted/index.html

Stanwood, J., (2020, October 20). *Violent incident on Miami flight caught on video*. Local10. https://www.local10.com/news/local/2020/10/20/violent-incident-on-miami-flight-caught-on-video/

Stephenson, G. M., Skinner, M., & Brotherton, C. J. (1976). Group participation and intergroup relations: An experimental study of negotiation groups. *European Journal of Social Psychology*, *6*(1), 51-70. https://doi.org/10.1002/ejsp.2420060105

Swann, W. B., Gómez, Á., Huici, C., Morales, J. F., & Hixon, J. G. (2010). Identity fusion and self-sacrifice: Arousal as a catalyst of pro-group fighting, dying, and helping behavior. *Journal of Personality and Social Psychology*, *99*(5), 824–841. https://doi.org/10.1037/a0020014

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science dducation. *Research in Science Education (Australasian Science Education Research Association)*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, *33*(1), 1-39. https://doi.org/10.1146/annurev.ps.33.020182.000245

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149-178. https://doi.org/10.1002/ejsp.2420010202

Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worschel (Eds.), *The social psychology of intergroup relations* (pp. 33- 47). Brooks/Cole Publishing.

TechTarget. (2021, December). *CCTV (closed circuit television)*. https://www.techtarget.com/whatis/definition/CCTV-closed-circuit-television

Thomas, K. (2018, June 8). *The murder of "Kitty" Genovese that led to the bystander effect & the 911 system*. The Vintage News. Retrieved October 12, 2021 from https://www.thevintagenews.com/2018/06/08/kitty-genovese/

Thornberg, R., Tenenbaum, L., Varjas, K., Meyers, J., Jungert, T., & Vanegas, G. (2012). Bystander motivation in bullying incidents: To intervene or not to intervene?. *The Western Journal of Emergency Medicine*, *13*(3), 247-252. https://doi.org/10.5811/westjem.2012.3.11792

Tilker, H. A. (1970). Socially responsible behavior as a function of observer responsibility and victim feedback. *Journal of Personality and Social Psychology*, *14*(2), 95–100. https://doi.org/10.1037/h0028773

Transportation, 49 U.S.C. § 46501 (2015). https://www.govinfo.gov/app/details/USCODE-2015-title49/USCODE-2015-title49-subtitleVII-partA-subpartiv-chap465-sec46501

Transportation Security Administration (2021). *Security directive 1542-21-01A, mask requirements*. https://www.tsa.gov/sites/default/files/sd-1542-21-01a.pdf

Tsang, S., Masiero, L., & Schuckert, M. (2018). Investigating air passengers' acceptance level of unruly in-flight behavior. *Tourism Analysis, 23*(1), 31–43. https://doi.org/10.3727/108354218X15143857349477

Tsaur, S.-H., Hsu, F.-S., & Kung, L.-H. (2020). Hassles of cabin crew: An exploratory study. *Journal of Air Transport Management, 85*. https://doi.org/10.1016/j.jairtraman.2020.101812

U.S. Census Bureau. (n.d.-a). United States: Populations and people. *2022 American Community Survey*. U.S. Department of Commerce. https://data.census.gov/profile/United_States?g=010XX00US#populations-and-people

U.S. Census Bureau. (n.d.-b). United States: Race and ethnicity. *2020 Decennial Survey*. U.S. Department of Commerce. https://data.census.gov/profile/United_States?g=010XX00US#populations-and-people

Van Cleemput, K., Vandebosch, H., & Pabian, S. (2014). Personal characteristics and contextual factors that determine "helping," "joining in," and "doing nothing" when witnessing cyberbullying: Bystander behavior in cyberbullying. *Aggressive Behavior*, *40*(5), 383–396. https://doi.org/10.1002/ab.21534

Van der Graaff, J., Branje, S., De Wied, M., Hawk, S., Van Lier, P., & Meeus, W. (2014). Perspective taking and empathic concern in adolescence: Gender differences in developmental changes. *Developmental Psychology, 50*(3), 881–888. https://doi.org/10.1037/a0034325

Van Griethuijsen, R. A. L. F., Van Eijck, M. W., Haste, H., den Brok, P. J., Skinner, N. C., Mansour, N., Gencer, A. S., & BouJaoude, S. (2014). Global patterns in students' views of science and interest in science. *Research in Science Education, 45*(4), 581-603. https://doi.org/10.1007/s11165-014-9438-6

Vanbeselaere, N. (1987). The effects of dichotomous and crossed social categorizations upon intergroup discrimination. *European Journal of Social Psychology, 17*(2), 143–156. https://doi.org/10.1002/ejsp.2420170203

Vredenburgh, A. N., Zackowitz, I. B., & Vredenburgh, A. G. (2015). Air rage: What factors influence airline passenger anger? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 59*(1), 400-404. https://doi.org/10.1177/1541931215591084

WABC. (2021, May 3). *Off-duty cop subdues woman after mid-air flight attendant attack on JFK flight*. Eyewitness News 7. https://abc7ny.com/chenasia-campbell-in-flight-fight-mid-flight-brawl-flight-attendant/10574832/

Wallace, G., & Muntean, P. (2021, July 28). *Flight attendants training in self-defense to deal with spike in unruly passengers*. CNN. https://www.wxii12.com/article/flight-attendants-self-defense-training-unruly-passengers/37159415

Wallach, M. A., Kogan, N., & Bem, D. J. (1964). Diffusion of responsibility and level of risk taking in groups. *Journal of Abnormal and Social Psychology*, *68*(3), 263–274. https://doi.org/10.1037/h0042190

Wang, X., French, B. F., & Clay, P. F. (2015). Convergent and discriminant validity with formative measurement: A mediator perspec*tive. Journal of Modern Applied Statistical Methods*, *14*(1), 83-106. https://doi.org/10.22237/jmasm/1430453400

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing, 27*(3), 236-247. https://doi.org/10.1016/j.ijresmar.2010.02.004

Weiner, B. (1980). May I borrow your class notes? An attributional analysis of judgments of help giving in an achievement-related context. *Journal of Educational Psychology, 72*(5), 676-681. https://doi.org/10.1037/0022-0663.72.5.676

Westland, C. J., (2010). Lower bounds on sample size in structural equation modeling. *Electronic Commerce Research and Applications, 9*(6), 476-487. https://doi.org/10.1016/j.elerap.2010.07.003

Whitney, I., & Smith, P. K. (1993). A survey of the nature and extent of bullying in junior/middle and secondary schools. *Educational Research, 35*(1), 3-25. https://doi.org/10.1080/0013188930350101

Whitley, A., & Gross, S. (2019, July 26). *Legroom on planes has been shrinking for years. It's about to get much, much worse*. Time. https://time.com/5636154/airplane-legroom-shrinking-asia/

Weiss, Y., & Fershtman, C. (1998). Social status and economic performance: A survey. *European Economic Review, 42*(3-5), 801-820. https://doi.org/10.1016/S0014-2921(97)00137-2

Wiesenthal, D. L., Austrom, D., & Silverman, I. (1983). Diffusion of responsibility in charitable donations. *Basic and Applied Social Psychology, 4*(1), 17-27. https://doi.org/10.1207/s15324834basp0401_2

Williams, C. R. (2000). A pain in the neck: Passenger abuse, flight attendants and emotional labour. *Journal of Occupational Health and Safety-Australia and New Zealand, 16*(5), 429-435.

Wilson, W., & Kayatani, M. (1968). Intergroup attitudes and strategies in games between opponents of the same or of a different race. *Journal of Personality and Social Psychology, 9*(1) 24-30. https://doi.org/10.1037/h0025720

Wilson, W., & Miller, N. (1961). Shifts in evaluations of participants following intergroup competition. *The Journal of Abnormal and Social Psychology, 63*(2), 428-431. https://doi.org/10.1037/h0043621

Wynd, C. A., Schmidt, B., & Schaefer, M. A. (2003). Two quantitative approaches for estimating content validity. *Western Journal of Nursing Research, 25*(5), 508–518. https://doi.org/10.1177/0193945903252998

Zhao, C. (2021, June 12). *Pilot Asks "All Strong Males" to Subdue "Problem Passenger" Who Turns Out to Be Flight Attendant*. Newsweek. https://www.newsweek.com/pilot-asks-all-strong-males-subdue-problem-passenger-who-turns-out-flight-attendant-1600112

Zoccola, P. M., Green, M. C., Karautsos, E., Katona, S. M., & Sabini, J. (2011). The embarrassed bystander: Embarrassability and the inhibition of helping. *Personality and Individual Differences*, *51*(8), 925–929. https://doi.org/10.1016/j.paid.2011.07.026

Appendix A

Permission to Conduct Research

Embry-Riddle Aeronautical University
Application for IRB Approval
EXEMPT Determination Form

Principal Investigator: Michael Pettit

Other Investigators: Scott R. Winter, Ph.D.

Role: Student **Campus:** Worldwide **College:** Aviation/Aeronautics

Project Title: Survey of Personality and Behavior

Review Board Use Only

Initial Reviewer: Teri Gabriel **Date:** 06/14/2023 **Approval #:** 23-137

Determination: Exempt

Dr. Shawn Doherty
IRB Chair Signature: Shawn M Doherty Digitally signed by Shawn M Doherty DN: cn=Shawn M Doherty, o=Embry-Riddle Aeronautical University, ou=Human Factors, email=dohertsh@erau.edu, c=US Date: 2023.06.19 14:01:48 -04'00'

Brief Description:

This research examines the applicability of an existing bystander intervention model to a flight attendant assault and attempts to identify factors which may affect a person's willingness to intervene. Participants will be asked to complete an online survey.

This research falls under the **EXEMPT** category as per 45 CFR 46.104:

☑ (2) Research that only includes interactions involving educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior (including visual or auditory recording) if at least one of the following criteria is met: (Applies to Subpart B [Pregnant Women, Human Fetuses and Neonates] and does not apply for Subpart C [Prisoners] except for research aimed at involving a broader subject population that only incidentally includes prisoners.)

Appendix B

Data Collection Device

D1 What is your age in years?

D2 Which of the following industries have you worked in? (Select all that apply)

- Financial (banking, investing, etc.)
- Advertising or marketing
- Construction
- Food (restaurant, catering, etc.)
- Commercial airline
- Automobile sales (new or used)
- Professional sports
- None of these

D3 What is your gender?

- Male
- Female

D4 When did you last use airline travel?

- Within the past year.
- Between 1 and 2 years ago.
- Between 2 and 3 years ago.
- More than 3 years ago.

D5 What is your highest education level?

- Less than high school
- High school graduate
- Some college (no degree)
- Associate's degree
- Bachelor's degree
- Master's degree

Doctorate degree

Skip

D6 What is your primary ethnicity?

Caucasian

African decent

Asian decent

Hispanic decent

Other

Skip

D7 What is your marital status?

Never been married

Married

Divorced

Separated

Widow or widower

Skip

PS1 If I saw someone being attacked (hit and pushed), my friends and parents would expect me to physically intervene.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

PS2 If I saw someone being attacked (hit and pushed), my friends and parents would expect me to tell the attacker to stop.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

PS3 If I saw someone being attacked (hit and pushed), my friends and parents would expect me to call for help.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

PS4 If I saw someone being attacked (hit and pushed), my friends and parents would think less of me if I minded my own business and did nothing.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

PS5 If I saw someone being attacked (hit and pushed), my friends and parents would think less of me if I left the area to avoid getting involved.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

IS1 I have law enforcement training.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

IS2 I have life-saving training.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

IS3 I have first aid training.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

IS4 I have self-defense training.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

IS5 I have medical training (doctor, nurse, LPN, etc.).

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

IS6 I have received training that can be used to help someone in need.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

SE1 I can always manage to solve difficult problems if I try hard enough.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

SE2 If someone opposes me, I can find means and ways to get what I want.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

SE3 It is easy for me to stick to my aims and accomplish my goals.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

SE4 I am confident that I could deal efficiently with unexpected events.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

SE5 Thanks to my resourcefulness, I know how to handle unforeseen situations.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

SE6 I can solve most problems if I invest the necessary effort.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

SE7 I can remain calm when facing difficulties because I can rely on my coping abilities.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

SE8 When I am confronted with a problem, I can usually find several solutions.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

SE9 If I am in trouble, I can usually think of something to do.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

SE10 No matter what comes my way, I'm usually able to handle it.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

Two strangers get into an argument, and one starts pushing and hitting the other person.

SO1 If I tried to physically make the aggressor stop, it would stop the argument.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

Two strangers get into an argument, and one starts pushing and hitting the other person.

SO2 If I told the aggressor to stop, it would stop the argument.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

Two strangers get into an argument, and one starts pushing and hitting the other person.

SO3 If I told others to make the aggressor stop, it would stop the argument.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

Two strangers get into an argument, and one starts pushing and hitting the other person.

VO1 If I tried to physically make the aggressor stop, it would make the victim feel better.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

Two strangers get into an argument, and one starts pushing and hitting the other person.

VO2 If I told the aggressor to stop, it would make the victim feel better.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

Two strangers get into an argument, and one starts pushing and hitting the other person.

VO3 If I told others to make the aggressor stop, it would make the victim feel better.

Strongly Disagree	Mostly Disagree	Undecided	Mostly Agree	Strongly Agree

Two strangers get into an argument, and one starts pushing and hitting the other person.

EO1 If I did something, it would help.

Strongly Disagree	Mostly Disagree	Undecided	Mostly Agree	Strongly Agree

SB1 Compared to most people, how honest are flight attendants?

A lot less Honest	A little less honest	About the same	A little more honest	A lot more honest

SB2 Compared to most people, how friendly are flight attendants?

A lot less friendly	A little less friendly	About the same	A little more friendly	A lot more friendly

SB3 Compared to most people, how hard working are flight attendants?

A lot less hard working	A little less hard working	About the same	A little more hard working	A lot more hard working

SB4 Compared to most people, how intelligent are flight attendants?

A lot less intelligent	A little less intelligent	About the same	A little more intelligent	A lot more intelligent

SB5 Compared to most people, how well-groomed are flight attendants?

A lot less well-groomed	A little less well-groomed	About the same	A little more well-groomed	A lot more well-groomed

SB6 Flight attendants have more admirable qualities than most people.

Strongly Disagree	Mostly Disagree	Undecided	Mostly Agree	Strongly Agree

You are sitting in an aisle seat on a mostly full airplane approximately halfway through your flight. You hear loud cursing and look up to see a passenger and a flight attendant, both women, standing in the aisle arguing a few rows away. You cannot tell what the argument is about, but you see the passenger begin trying to punch the flight attendant as she continues to yell at her.

NC1 If I do nothing, this incident puts me at risk of getting seriously hurt.

Strongly Disagree	Mostly Disagree	Undecided	Mostly Agree	Strongly Agree

NC2 If I do nothing, this incident could end up costing me money.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

NC3 If I do nothing, this incident could end up disrupting my travel plans.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

NC4 If I do nothing, this incident could make me feel anxiety, scared, stressed, etc.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

NC5 If I do nothing, this incident could negatively affect me.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

RN1 It is evident to me that the flight attendant urgently needs help.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

RN2 I believe in this situation, passengers should get involved.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

RN3 I believe that this situation is an emergency that requires help from passengers.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

TR1 It is my responsibility to personally intervene and assist the flight attendant.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

TR2 Even though I'm not involved, it is still my responsibility to try and stop it.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

TR3 If I get involved, my actions will help reduce these kinds of incidents.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

KW1 I know what to say to get others to help the flight attendant.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

KW2 I know what to say to get the person to stop.

Strongly Disagree | Mostly Disagree | Undecided | Mostly Agree | Strongly Agree

KW3 I know how to physically help the flight attendant.

Strongly Disagree	Mostly Disagree	Undecided	Mostly Agree	Strongly Agree

DA1 I would intervene even if I am not sure other passengers would support me.

Strongly Disagree	Mostly Disagree	Undecided	Mostly Agree	Strongly Agree

DA2 I would intervene even though I might look foolish.

Strongly Disagree	Mostly Disagree	Undecided	Mostly Agree	Strongly Agree

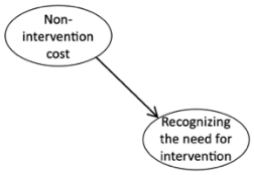
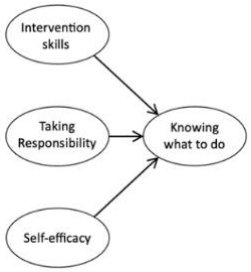
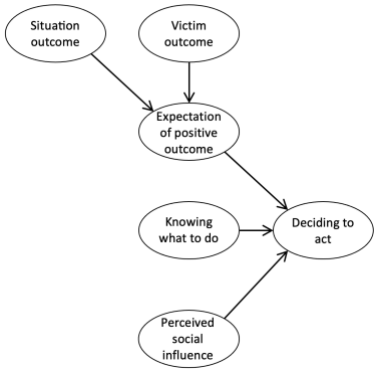
Z1 Are there any comments you would like to share about either the survey or the scenario?

Appendix C

Tables

Table C1

Comparison of Sample Size Methods for Partitioned and Full Model Relationships

Partition	Rule of 10	Minimum R^2	Ratio method	Inverse square root	Paired latent variables method
	30	110	200	241	359
	50	124	100	241	550
	50	124	116	241	550
	30	124	100	241	550

Partition	Rule of 10	Minimum R^2	Ratio method	Inverse square root	Paired latent variables method
Full model (all constructs, predictors, and relationships)	50	124	88	241	765

Table C2

Original and Modified Indicator Context

Indicator	Original Context	Modified Context
PS1	If in my classroom someone repeatedly bullies another classmate, according to my classmates I should intervene to help victim.	If I saw someone being attacked (hit and pushed), my friends and parents would expect me to physically intervene.
PS2		If I saw someone being attacked (hit and pushed), my friends and parents would expect me to tell the attacker to stop.[a]
PS3	If in my classroom someone repeatedly bullies another classmate, according to my classmates I should apprise an adult of what is happening so that he can intervene.	If I saw someone being attacked (hit and pushed), my friends and parents would expect me to call for help.
PS4	If in my classroom someone repeatedly bullies another classmate, according to my classmates I should do nothing because it's not my business.	If I saw someone being attacked (hit and pushed), my friends and parents would think less of me if I minded my own business and did nothing.

Indicator	Original Context	Modified Context
PS5	If in my classroom someone repeatedly bullies another classmate, according to my classmates I should withdrawal for self-protection.	If I saw someone being attacked (hit and pushed), my friends and parents would think less of me if I left the area to avoid getting involved.
IS1	Do you have police training?	I have law enforcement training.
IS2	Do you have life-saving training?	I have life-saving training.
IS3	Do you have first aid training?	I have first aid training.
IS4	Do you have self-defense training?	I have self-defense training.
IS5	Do you have medical training?	I have medical training (doctor, nurse, LPN, etc.).
IS6		I have received training that can be used to help someone in need.
SE1	I can always manage to solve difficult problems if I try hard enough.	I can always manage to solve difficult problems if I try hard enough.

Indicator	Original Context	Modified Context
SE2	If someone opposes me, I can find means and ways to get what I want.	If someone opposes me, I can find means and ways to get what I want.
SE3	It is easy for me to stick to my aims and accomplish my goals.	It is easy for me to stick to my aims and accomplish my goals.
SE4	I am confident that I could deal efficiently with unexpected events.	I am confident that I could deal efficiently with unexpected events.
SE5	Thanks to my resourcefulness, I know how to handle unforeseen situations.	Thanks to my resourcefulness, I know how to handle unforeseen situations.
SE6	I can solve most problems if I invest the necessary effort.	I can solve most problems if I invest the necessary effort.
SE7	I can remain calm when facing difficulties because I can rely on my coping abilities.	I can remain calm when facing difficulties because I can rely on my coping abilities.
SE8	When I am confronted with a problem, I can usually find several solutions.	When I am confronted with a problem, I can usually find several solutions.

Indicator	Original Context	Modified Context
SE9	If I am in trouble, I can usually think of something to do.	If I am in trouble, I can usually think of something to do.
SE10	No matter what comes my way, I'm usually able to handle it.	No matter what comes my way, I'm usually able to handle it.
SO1	If you tried to make the others stop the bullying, it would decrease or stop the bullying.	If I tried to physically make the aggressor stop, it would stop the argument.[b]
SO2	If you tried to comfort the victim or encourage him/her to tell the teacher about bullying, it would stop the bullying.	If I told the aggressor to stop, it would stop the argument.[b]
SO3	If you tried to tell others to stop bullying or say that bullying is stupid, it would stop the bullying.	If I told others to make the aggressor stop, it would stop the argument.[b]
VO1	If you tried to make the others stop the bullying, it would make the bullied person feel better.	If I tried to physically make the aggressor stop, it would make the victim feel better.[b]

Indicator	Original Context	Modified Context
VO2	If you tried to comfort the victim or encourage him/her to tell the teacher about bullying, it would make the bullied person feel better.	If I told the aggressor to stop, it would make the victim feel better.[b]
VO3	If you tried to tell others to stop bullying or say that bullying is stupid, it would make the bullied person feel better.	If I told others to make the aggressor stop, it would make the victim feel better.[b]
EO1		If I did something, it would help.
SB1	How many immigrants do you think are honest?	Compared to most people, how honest are flight attendants?
SB2	How many immigrants do you think are friendly?	Compared to most people, how friendly are flight attendants?
SB3	How many immigrants do you think are hardworking?	Compared to most people, how hard working are flight attendants?
SB4	How many immigrants do you think are unintelligent?	Compared to most people, how intelligent are flight attendants?

Indicator	Original Context	Modified Context
SB5	How many immigrants do you think are clean?	Compared to most people, how well-groomed are flight attendants?
SB6		Flight attendants have more admirable qualities than most people.
NC1		If I do nothing, this incident puts me at risk of getting seriously hurt.
NC2		If I do nothing, this incident could end up costing me money.
NC3		If I do nothing, this incident could end up disrupting my travel plans.
NC4		If I do nothing, this incident could make me feel anxiety, scared, stressed, etc.
NC5		If I do nothing, this incident could negatively affect me.

Indicator	Original Context	Modified Context
RN1	It is evident to me that urgent humanitarian aid is needed for the Syrian refugee issue.	It is evident to me that the flight attendant urgently needs help.
RN2	I think that the Syrian refugee issue is a severe emergency that other people should be involved.	I believe in this situation, passengers should get involved.
RN3	I believe that the situation of Syrian refugees is an emergency that requires the help of other people.	I believe that this situation is an emergency that requires help from passengers.
TR1	I feel personally responsible to intervene and assist in resolving bullying or sexual harassment incidents.	It is my responsibility to personally intervene and assist the flight attendant.
TR2	If I am not the one bullying or harassing others, it is still my responsibility to try to stop it.	Even though I'm not involved, it is still my responsibility to try and stop it.
TR3	I believe that my actions can help to reduce bullying and sexual harassment	If I get involved, my actions will help reduce these kinds of incidents.

Indicator	Original Context	Modified Context
KW1	I know what to say to get others to help or support Syrian refugees.	I know what to say to get others to help the flight attendant.
KW2	I can find an organization that provide support to Syrian refugees.	I know what to say to get the person to stop.
KW3	I know a number of ways I can help Syrian refugees.	I know how to physically help the flight attendant.
DA1	I am hesitant to intervene when a man's sexual conduct is questionable because I am not sure other people would support me.	I would intervene even if I am not sure other passengers would support me.
DA2	Even if I thought it was my responsibility to intervene to prevent an assault, I might not out of a concern I would look foolish.	I would intervene even though I might look foolish.

Note. Original verbiage is from the following sources: PS (Pozzoli & Gini, 2010), IS (Huston et al., 1981), SE (Schwarzer, 1999), SO (Pöyhönen & Salmivalli, 2012), VO (Pöyhönen & Salmivalli, 2012), SB (Abbott & Cameron, 2014), RN (Albayrak-Aydemir & Gleibs, 2021), TR (Nickerson et al., 2014), KW (Albayrak-Aydemir & Gleibs, 2021), DA (Burn, 2009). *Nonintervention cost* (NC) is an original measure

[a] PS2 captures direct verbal intervention. This dimension was not included in the original context, but is included to better capture all dimensions

[b] The preceding sentence for SO and VO indicators is “Two strangers get into an argument, and one starts pushing and hitting the other person.” The sentence remained displayed for SO and VO questions.

Appendix D

Figures

Figure D1

SME Assessment of Nonintervention Cost Content Validity

Consider the following scenario:

You are sitting in an aisle seat on a mostly full airplane approximately halfway through your flight. You hear loud cursing and look up to see a passenger and a flight attendant, both women, standing in the aisle arguing a few rows away. You cannot tell what the argument is about, but you see the passenger begin trying to punch the flight attendant as she continues to yell at her.

"Nonintervention cost": Potential negative impacts to a passenger if nobody helps.

Survey Question: "If I do nothing, this incident could end up disrupting my travel plans."
Is this part of the construct "Nonintervention cost"?

Definitely NOT a part	Mostly NOT a part	Just barely NOT a part	Just barely IS a part	Mostly IS a part	Definitely IS a part

Survey Question: "If I do nothing, this incident could end up costing me money."
Is this part of the construct "Nonintervention cost"?

Definitely NOT a part	Mostly NOT a part	Just barely NOT a part	Just barely IS a part	Mostly IS a part	Definitely IS a part

Survey Question: "If I do nothing, this incident puts me at risk of getting seriously hurt."
Is this part of the construct "Nonintervention cost"?

Definitely NOT a part	Mostly NOT a part	Just barely NOT a part	Just barely IS a part	Mostly IS a part	Definitely IS a part

Survey Question: "If I do nothing, this incident could make me feel anxiety, scared, stressed, etc."
Is this part of the construct "Nonintervention cost"?

Definitely NOT a part	Mostly NOT a part	Just barely NOT a part	Just barely IS a part	Mostly IS a part	Definitely IS a part

Are there any parts of nonintervention cost not identified?