UiT The Arctic University of Norway

The Faculty of Health Sciences

Department of Psychology

Shaping the Perception of Risk

Investigating a Paradigmatic Case of Applied Decision Making under Uncertainty

Matthew B. Stephensen

A dissertation for the degree of Philosophiae Doctor – June 2021

Table of Contents

List of Tables

List of Figures

Acknowledgments

I am tremendously grateful for the guidance and support of my main supervisor Torsten Martiny-Huenger and my co-supervisors Christin Schulze and Audun Hetland. Thank you for all that you helped me learn and achieve.

Special thanks go to my wife Beatrice and my children Zoe, Sofia, and Liam who supported and endured me during this quixotic pursuit. This would not have been possible or worthwhile without you.

Abstract

Risk and uncertainty characterise most human endeavours. Whether judging risk when buying a house, making a career change, considering marriage, or simply ordering food at an unfamiliar restaurant, each decision involves evaluating the available evidence to infer the likelihood of negative outcomes. Such decisions are often marked by uncertainty, the condition of incomplete knowledge about the likelihoods of possible outcomes, which can originate from the unpredictability of the environment or the cognitive limitations of the decision maker. This thesis aims to advance our understanding of how people routinely grapple with complex risk judgments and decisions under uncertainty. Using the exemplary case of decision making in avalanche terrain, this thesis investigates three specific aspects of how a population with domain-specific experience (backcountry skiers) perform a complex decision task (judging avalanche risk and deciding whether to ski) in a highly uncertain context (avalanche terrain). Paper 1 tests how the qualitative frame of the question eliciting a risk judgment influences risk perception and behaviour intention, and infers the cognitive process that the question frame activates. Risk judgments framed in terms of safety were found to result in more cautious, conservative judgments than when framed in terms of danger. That effect suggests that the question frame directed attention in a way that guided selective information sampling. Paper 2 investigates the relation between affective evaluations and risk judgments across multiple judgments of specific instances of an activity distinguished by contextual information. The perceived safety of a scenario was found to positively relate to the degree of liking that scenario across multiple judgments of specific instances of an activity. That provides a dynamic picture of the degree to which cognitive and affective factors are intertwined. Paper 3 examines the roles of uncertainty, domain-specific experience, and affective evaluations in the formation of confidence in complex risk judgments. Domain-specific experience was found to magnify confidence. However, greater experience does not necessarily correspond to more accurate risk judgments or better calibrated confidence. Furthermore, affective evaluations were found to play a role in the formation of confidence. Confidence was higher when the decision maker's affective evaluation matched their risk perception, and less confident when there was a mismatch between the two. The findings are discussed within the wider framework of studying the decision process in order to help people make better decisions. This research has theoretical and applied relevance for backcountry skiing in avalanche terrain and more broadly in other fields of applied risk judgments and decisions.

List of Papers

Paper I

Stephensen, M. B., Schulze, C., Landrø, M., Hendrikx, J., & Hetland, A. (2021). Should I judge safety or danger? Perceived risk depends on the question frame. *Journal of Experimental Psychology: Applied*. https://doi.org/10.1037/xap0000354

Paper II

Stephensen, M. B., & Martiny-Huenger, T. (2021). Liking and perceived safety across judgments of distinct instances of a category of activity. *Journal of Risk Research*. https://doi.org10.1080/13669877.2021.1905693

Paper III

Stephensen, M. B., Martiny-Huenger, T., Schulze, C. (2021). Confidence in complex risk judgments: the roles of uncertainty, experience, and affect. *Preprint currently under peer-review*.

1 Introduction

Studying applied risk judgments and decisions can be a frustratingly messy affair. The complex, noisy nature of most decision environments and the multiplicity of cognitive processes potentially activated during the decision process make the study of applied decision making a daunting task. Yet investigating how a population judges risk and decides behaviour in specific information environments is essential for advancing the science of judgement and decision-making (Fischhoff & Broomell, 2020). The characteristics of a decision task, the nature of its information environment, and the decision maker's domain-specific experience can each influence the decision making process. It is therefore important that the design of experiments used in applied research represent the multifaceted characteristics of the real-life conditions to which experimental findings are generalized (Brunswik, 1944, 1955a; Dhami et al., 2004; Hammond, 1966; Juslin & Montgomery, 2007). Research on applied decision making should involve the complex, goal-directed judgements and decisions that typify human behaviour rather than focus exclusively on simple, one-dimensional tasks with well-defined parameters of information (Dhami et al., 2004; Juslin, 1994; Juslin et al., 2000; Yeung & Summerfield, 2012). Accordingly, this thesis employs the task of judging avalanche risk and deciding whether to ski to study applied risk judgments and decisions under uncertainty.

The simplicity of skiing in the mountains is almost comic. First, go to a snow-covered mountain. Strap a plank to each foot, take a stick in each hand, and walk up that snow-covered mountain. Finally, let gravity propel your thrilling descent down the mountain – joyous experience that can, at times, almost feel like flying. Yet the beauty and simple pleasure of skiing belie its true complexity, uncertainty, and risk. The combination of snow and mountains required for skiing is the very combination that creates the necessary conditions for snow avalanches. Avalanche terrain is a complex, variable, and uncertain environment in which a decision error can result in severe injury or death. For the increasing number of people venturing into the backcountry to ski (Birkeland et al, 2017), their safety depends on their ability to understand, identify, and mitigate risk in a hazardous, uncertain environment.

Judging risk involves evaluating the available evidence to make inferences about the likelihood of negative outcomes and their potential impact. For backcountry skiers, that requires evaluating macro-level cues such as the regional avalanche danger forecast and the weather history, and micro-level cues such as slope angle, terrain features, and snowpack characteristics to infer the stability of a slope and the likelihood of triggering an avalanche if they decide to

ski. An incorrect decision to ski an unstable slope can result in an avalanche that can injure or kill those in its path. Unfortunately, judging avalanche risk is often highly uncertain. The complexity of the task, and the variability and inherent unpredictability of the decision environment often create a state of incomplete knowledge about the likelihood of those negative outcomes (Engeset et al., 2018; Furman et al., 2010; Landrø et al., 2020a, 2020b; Statham et al., 2018). As a result, decision errors resulting in avalanche accidents and fatalities are not uncommon.

The Center for Avalanche Research and Education (CARE) at UiT The Arctic University of Norway studies how backcountry skiers make risk judgments and decisions within the complex environment of avalanche terrain with the aim of improving decision making practices to ultimately decrease the frequency of fatal errors. As a CARE-funded doctoral research fellow, my thesis investigates how experienced backcountry skiers respond to the task of judging avalanche risk. However, the relevance of my thesis is not limited to that specific context or task. Decision making in avalanche terrain offers an exemplary opportunity to study the complex risk judgments and decisions of an experienced, highly motivated population under authentic levels of uncertainty. The findings from this research can be generalized to a range of applied disciplines in which people search for and evaluate the available evidence to make inferences about the likelihood of negative outcomes and then decide the most appropriate behaviour. Analysing specific decision tasks and the processes by which people make those decisions increases the heterogeneity of the methods used and the populations and decision tasks investigated in the decision sciences, thereby expanding the knowledge base and relevance of the field (Fischhoff & Broomell, 2020). Ultimately, my research uses the uncertain decision environment of avalanche terrain and the complex task of judging avalanche risk to advance our understanding of how people routinely grapple with complex decisions under high uncertainty.

For the present thesis, I use hypothetical scenarios of an applied decision context (skiing in avalanche terrain) and a population with domain-specific experience (backcountry skiers) to examine three distinct aspects of the risk judgment and decision process (judging avalanche risk and deciding whether to ski). The study of each of those three aspects of the decision process is presented in one of three empirical papers. In paper 1, we investigate the effect of question framing on risk perception and behaviour intention. In paper 2, we examine the relation between affective evaluations and risk judgments (i.e., the affect heuristic). In paper 3,

we explore the formation of confidence in complex risk judgments. I consider those three topics in greater detail below when presenting the objectives of this thesis.

The purpose of what follows is to contextualize and complement those three empirical papers rather than merely summarize them. I first elaborate upon the virtues and benefits of studying risk judgments and decisions in the context of decision making in avalanche terrain. In response to the imagined question, "*Why study decision making in avalanche terrain?"* I consider how that context involves a meaningful, goal-oriented decision task for an experienced and engaged population, and the benefits that that affords my research. I then describe the characteristics of the decision task and its information environment, and consider how decision making in avalanche terrain represents a paradigmatic case for studying complex risk judgments and decisions under high uncertainty. Having described the applied context that frames my thesis, I then present the objective of my research (Section 2). I introduce the three aspects of the risk judgment and decision process that my research examines in the three papers, discuss their relevance to the decision task, and define the research questions that we attempt to answer in each paper. After that introduction to the context and decision task employed in my research, I explain the methodological approach used throughout this project (Section 3) before providing a comparative overview of the research conducted for each of the three papers (Section 4). Finally, I summarize their results (Section 5) and broadly consider the implications of my research findings beyond what is discussed in each of the three papers (Section 6).

1.1 Judging Avalanche Risk: A Meaningful, Goal-Oriented Decision Task for an Experienced Population

Judging avalanche risk offers an exemplary opportunity to study complex, applied decisions. However, there are numerous ethical concerns with sending research participants into avalanche terrain under conditions of high uncertainty to study their risk judgments and decisions. My co-authors and I therefore used hypothetical backcountry skiing scenarios throughout this research. We modelled those scenarios on the natural decision environment using historic reports of avalanche terrain from national and regional avalanche authorities and the technical guidance of avalanche experts. Using hypothetical scenarios as experimental stimuli also solved the logistical challenge of aligning the conditions needed to create the desired level of uncertainty with the availability of participants. Fortunately, with a

comprehensive understanding of avalanche terrain it is possible to design realistic hypothetical scenarios that reflect the information properties and affective richness of the environment in which those tasks are naturally embedded (Hammond, 1966). Backcountry skiers evaluate a multiplicity of information cues to judge the risk of avalanches. CARE researchers (Landrø et al., 2020a, 2020b) investigating the information cues used in various decision making frameworks identified 53 different information cues that are used for avalanche risk judgments. Those information cues include macro-level cues – regional cues that must be interpreted for local application as and when relevant – such as the forecasted avalanche danger level, the forecasted avalanche problems, and weather history; and micro-level cues – context-specific information cues – such as slope angle, terrain features, and snowpack characteristics. All of those information cues can be communicated through text and images, and their configuration can be manipulated in hypothetical scenarios to create different degrees of risk and uncertainty.

The information cues we used to design the hypothetical scenarios represent the very information that backcountry skiers evaluate to judge avalanche risk both before and during tours in avalanche terrain. By maintaining their natural richness, complexity, and uncertainty, we aimed for the stimuli presented to participants to be authentic representations of the contexts in which the decision task is naturally situated (Dhami et al., 2004). In such scenarios, the decision tasks remain entangled with the multiple contextual variables that are present in the decision environment (Morrow, 2018), many of which might be irrelevant for the decision task. Importantly, the activity, environment, and decision task depicted in the scenarios are familiar and meaningful to a population of backcountry skiers. By designing stimuli that reflect the richness and complexity of avalanche terrain, we sought to achieve the realism needed to engage their natural decision processes (Dhami et al., 2004; Fischhoff & Broomell, 2020). This enables an examination of the intersection between the information environment and the decision process (Hertwig et al., 2019) and increases the external validity and generalizability of the research findings (Dhami et al., 2004). In the methods section further below (Section 3), I explain the methodological approach we used to design realistic hypothetical scenarios that reflected the complex, rich information environment in which the decision task is naturally situated. I also consider the limitations of that approach.

Our use of the applied context of decision making in avalanche terrain also made it possible to recruit participants with domain-specific experience (i.e., backcountry skiers) to participate in our studies. This overcomes some of the problems associated with the use of student samples such as the limits when generalizing from students to the general public (Hanel & Vione, 2016).

Judging avalanche risk and deciding whether to ski under specific conditions are not trivial tasks for backcountry skiers. That the task of judging avalanche risk, the activity of backcountry skiing, and the environment of avalanche terrain presented in the experimental stimuli are familiar, meaningful, and important for the participants increases the ecological validity of our findings (Brunswik, 1944, 1955a; Dhami et al., 2004; Hammond, 1966; Juslin & Montgomery, 2007). In addition, backcountry skiers are invested in research on decision making in avalanche terrain because the findings have implications for their own pleasure and wellbeing. We were therefore able to recruit large numbers of backcountry skiers to participate in numerous experimental studies (the value of being able to recruit large samples of participants with domain-specific experience should not be underestimated). Decision making in avalanche terrain offers an excellent opportunity to study authentic decision tasks performed by authentic participants who are knowledgeable about those tasks and their environment, and for whom the task and environment have affective value and behaviour-relevant, goal-oriented meaning (Morrow, 2018).

1.2 Judging Avalanche Risk: A Paradigmatic Case of Risk Judgments and Decisions under Severe Uncertainty

Avalanche terrain is any mountainous area where the combination of terrain, snow, and weather can create the necessary conditions for a snow avalanche. It is a complex and highly uncertain environment that is objectively dangerous under a range of conditions (Engeset et al., 2018; Landrø et al., 2020a, 202b; Niedermeier et al., 2020; Statham et al., 2018). An avalanche occurs when the snowpack fails due to instability, causing a mass of snow to suddenly and violently slide down the slope at high velocity. Human activities such as skiing can cause the failure of the snowpack triggering the release of an avalanche. Although avalanche accidents are unintended and unexpected, they are not random or unexplainable; they require that the victims exposed themselves to the risk of avalanche (Johnson et al., 2020). To manage that risk, backcountry skiers must search for and evaluate a multiplicity of information cues in a complex, uncertain, and highly variable environment to make predictive inferences about the likelihood of an avalanche and its potential impact. Yet that is no simple task. In 90% of fatal avalanche accidents, the victim or someone in the victim's party triggered the avalanche (McClung & Schaerer, 2006; Schweizer & Lütschg, 2000). That statistic implies that accidents occur when people mistakenly perceive objectively dangerous conditions as safe. People's risk judgments

and decisions are critical factors in avalanche accidents and fatalities. It is therefore essential to consider how people judge risk and to examine the factors that underlie the decision process (Slovic, 1987).

Decision making in avalanche terrain is an ideal task for investigating risk judgments and decisions under high uncertainty. Avalanche terrain is a complex decision environment where no single cue or combination of cues allows the perfect prediction of risk (Landrø et al., 2020a, 2020b). The validity and reliability of the available evidence are often limited, ambiguous, or conflicting. Ascertaining when the combination of environmental factors such as terrain, slope angle, weather history, and snowpack characteristics present a real danger to skiers and deciding whether to ski are extremely complex tasks. The complexity, unpredictability, and spatial and temporal variability of the decision environment often creates a state of incomplete knowledge about the probabilities of possible outcomes (Hertwig et al., 2019; Speekenbrink & Shanks, 2013; Statham et al., 2018). Such stochastic nature and inherent unpredictability of the environment are a source of considerable environmental uncertainty (Fox & Ülkümen, 2011; Gillies, 2000; Kozyreva & Hertwig, 2021; Ülkümen et al., 2016). Consequently, it is virtually impossible to eliminate environmental uncertainty when assessing avalanche risk (Engeset et al., 2018; Furman et al., 2010; Landrø et al., 2020a, 2020b; Statham et al., 2018).

The internal uncertainty of the decision maker is also highly prevalent during avalanche risk judgments and decisions. Because of the complexity of the task, uncertainty can stem from the limited knowledge and skill of even the most experienced backcountry skiers. A task-specific strategy defines how a decision maker approaches a complex decision task such as judging avalanche risk (Rieskamp & Otto, 2006). The accuracy of backcountry skier's decision, and by extension the accuracy of their decision strategy, is the degree to which it corresponds with the true state of the world (Brunswik, 1955b; Fischhoff & Broomell, 2020; Hammond et al., 1975). This is in contrast to measuring the quality of a decision according to the alternative criteria of coherence, which concerns reasoning in accordance with the rules of logic and probability theory in a consistent way (Fischhoff & Broomell, 2020; Hammond, 1996). That avalanche accidents are overwhelmingly the result of human error indicates that accidents occur when the perceived avalanche risk fails to correspond with environmental reality. A broader conceptualization of this benchmark of accuracy adopts a consequentialist interpretation of rationality in cognition (Kozyreva & Hertwig, 2021; Schurz & Hertwig, 2019) and measures accuracy in terms of the success in achieving one's goal in the world. The measure of accuracy concerns the degree to which a decision corresponds with reality so that the decision maker

achieves their goal under the constraints and affordances of the decision environment and the capacity of the decision maker (Kozyreva & Hertwig, 2021). In avalanche terrain, this goal would be *to not trigger an avalanche*. Irrespective of whether the correspondence or consequentialist perspective on decision accuracy is employed, the most common measure of decision accuracy is outcome feedback: whether or not a skier triggered an avalanche.

Backcountry skiers infer the accuracy of their performance based on the outcome of their decisions. Unfortunately, avalanche terrain is an environment where poor decisions are frequently marked by invalid feedback (Ebert, 2019; Johnson et al., 2020; Zweifel & Haegeli, 2014). Although an avalanche event is a definitive indication of decision error, not triggering an avalanche is not a definitive indication of a correct decision. It is estimated that the majority of mistaken decisions in avalanche terrain fortuitously *do not* result in an avalanche; luck rather than decision accuracy is the cause of many positive outcomes (Tremper, 2008). That implies that outcome feedback is an unreliable measure of decision accuracy in avalanche terrain. This unreliability of outcome feedback and the absence of representative performance feedback make it problematic for decision makers to evaluate the accuracy of their decisions, refine their decision strategy, and calibrate their level of competency. Such a wicked learning environment obstructs skill development and magnifies the likelihood that decision makers have incomplete individual knowledge about the probabilities of possible outcomes (Hogarth et al., 2015). This can result in overconfidence and the proliferation of erroneous practices in the decision strategy. Decision making in avalanche terrain represents a context in which greater experience might not necessarily correspond to greater expertise. A mismatch between perceived risk and reality could be present for even the most experienced backcountry skiers in avalanche terrain. Due to this combination of external, environmental uncertainty and internal, cognitive uncertainty, decision making in avalanche terrain represents a paradigmatic applied case of complex risk judgements and decisions under high uncertainty.

2 Research Objectives

The objective of this thesis was to increase our understanding of the cognitive processes that underlie complex risk judgments and decisions under uncertainty. Given the characteristics of the decision task and information environment that I described in the previous section, decision making in avalanche terrain offered an exemplary context for conducting research in pursuit of that objective. However, because a multiplicity of cognitive processes are potentially activated during such a complex task, I focused my research on the cognitive processes associated with three distinct aspects of decision making in avalanche terrain. The study of each of those three aspects of the decision process is presented in one of three empirical papers:

1) In paper 1, my co-authors and I test how the qualitative frame of the question eliciting a risk judgment influences risk perception and behaviour intention. Based on the measured effect, we infer the cognitive process that the question frame activates.
2) In paper 2, we investigate the relation between affective evaluations and risk judgments (i.e., the affect heuristic) across multiple judgments of specific instances of an activity distinguished by contextual information.
3) In paper 3, we examine the roles of uncertainty, domain-specific experience, and affective evaluations in the formation of confidence in complex risk judgments

We chose those three topics because of their apparent prominence in the process of judging avalanche risk and deciding whether to ski, but also because of their theoretical relevance to the field of applied judgments and decision making more broadly. There were knowledge gaps within the existing literature on framing effects, affective evaluations, and confidence that we believed could be effectively studied using this applied context of judgment and decision making. Question framing, affective evaluations, and retrospective confidence are important natural elements of the decision making in avalanche terrain. Moreover, they are topics that we can effectively study using the methods that we envisioned. Studying them in the context of risk judgments and decisions in avalanche terrain held the promise of increasing our understanding of the cognitive processes that drive the decisions in that specific context and in other similar contexts of high uncertainty.

A potential secondary outcome from this research is that by better understanding the cognitive processes that underlie risk judgments and decisions under uncertainty, we might identify

opportunities to harness, influence, or mitigate those processes to help decision makers do better. Such an outcome would identify important avenues for future research. This secondary outcome is particularly relevant for the CARE's research programme and beyond to other fields of applied risk perception and communication. Although Paper 1 considers the possibility of harnessing question framing to increase the likelihood of more cautious, conservative behaviour in the face of uncertainty, Papers 2 and 3 do not draw similar conclusions about affective evaluations or confidence. I therefore consider this secondary outcome of "better risk judgments and decisions" later in the General Discussion (Section 6) when considering the results and implications of the three papers.

In the following sub-sections, I introduce the research questions that we addressed in each of the three papers and summarize their findings. As each paper studies a different aspect of the decision making process, the theoretical context of each paper is unique. In each paper, we situate the scope and results of that research within the wider context of the current theory and empirical findings for that topic. I therefore refrain from unnecessarily repeating that contextualization here, and only briefly introduce the theoretical and applied importance of each paper at this time.

2.1 Paper 1: Risk Perception and Question Framing

Research has demonstrated that framing influences how people respond to or evaluate an object, activity, or situation (for reviews see e.g., Kühberger, 1998, 2017; Levin et al., 1998; Maule & Villejoubert, 2007; Piñon & Gambara, 2005). Research on framing effects informs policies and practices in applied domains such as health (e.g., Garcia-Retamero & Cokely, 2011; Garcia-Retamero & Galesic, 2010; Peters et al., 2011; Rothman & Salovey, 1997) and finance (Kirchler et al, 2005; Weber et al., 2000). Although framing can have a troublesome biasing effect in many circumstances, not all instances of framing are considered harmful or problematic. The purposeful use of framing to promote specific judgments and decisions is an established practice in several fields such as media and communications (Block & Keller, 1995; Chong & Druckman, 2007; Scheufele & Tewksbury, 2007) and marketing (Biswas, 2009; Biswas & Grau, 2008; Levin & Gaeth, 1988). Might framing be strategically employed to increase the likelihood of more cautious, conservative judgments and decisions in avalanche terrain?

When judging risk, we evaluate an attribute of an object, action, or situation. Attributes are commonly conceptualized in terms of polarized dimensions. For example, we understand the attribute speed in terms of the two dimensions "fast" and "slow". Those two dimensions are like the two sides of a coin: distinct but inseparable. The two dimensions provide opposing but complementary perspectives on the object of judgment. When judging an attribute such as risk, it is a natural process of language to focus on a single dimension of that judged attribute (Hilton, 2011). For example, when judging speed, we commonly ask "Is it *fast*?" or, alternatively, "Is it *slow*?" rather than formulating a judgment using both dimensions. Similarly, it is natural for backcountry skiers to formulate the judgment as either "How safe are the conditions?" or "How dangerous are the conditions?" Crucially, the question frame defines a decision maker's perspective on the target of judgment and establishes the reference point against which they make their judgment.

The use of framing in applied contexts depends on correctly matching the type of frame – and the cognitive process that it activates – with the objective for its application. The cognitive processes that a frame activates and the direction of the resulting framing effect likely depend on the form and domain in which the frame is achieved. Consequently, the different theoretical accounts of the framing effect – the valence account, the communication account, and the attention account – make conflicting predictions about the direction and prevalence of an effect evoked by a question frame. Framing effects have been studied extensively in the context of framing the presentation of information about the object(s) of judgment. However, uncertainty persists about the cognitive processes (and thus the direction of any possible effect) that are activated by the qualitative attribute that frames the question eliciting a judgment (i.e., question framing). Previous research on question framing produced conflicting results on the direction of the framing effect (e.g., Payne et al., 2013; Comerford & Robinson, 2017). We therefore sought to determine how the choice of frame influences perceived risk and behavioural decisions and, based on that effect, to infer the cognitive process that a question frames activates.

In Paper 1, we investigated how the qualitative frame of the question that elicits a risk judgment influences the perception of avalanche risk and the decision (intention, not action) of whether to ski. Risk judgments framed in terms of safety (*How safe is it?*) were found to result in more cautious, conservative judgments than when framed in terms of danger (*How dangerous is it?*). The direction of that effect suggests that the question frame directed attention in a way that guided selective information sampling. The question frame was also found to have an indirect

influence on behaviour intention via the direct effect on the risk judgment. Framing the judgment of avalanche risk in terms of safety increased the likelihood of more cautious, conservative risk perception and behaviour intention.

2.2 Paper 2: Affective Evaluations and Perceived Risk

Affective evaluations can serve as a compelling orienting mechanism for a range of decision tasks including risk judgments (Blanchette & Richards, 2010; Finucane et al, 2000; Lerner et al., 2015; Lowenstein et al., 2001; Schwarz, 2007, 2012; Slovic, 1987; Slovic et al., 2002, 2004). The more a person likes an object or activity, the safer and more valuable it appears, whereas the more a person dislikes an object or activity, the more unsafe and less valuable it appears (Alhakami & Slovic, 1994; Finucane et al., 2000; Slovic et al., 2002). That relation is particularly important in contexts of high uncertainty such as avalanche terrain. When valid, definitive evidence of risk is lacking and the available evidence is ambiguous or conflicting, reliance upon affective evaluations can increase and might even be considered justified, as the terms “gut instinct” and “intuition” suggest (Wilson & Arvai, 2006).

Evidence for this positive relation between liking and perceived safety (or disliking and perceived danger) was found when judging categorical representations of objects and activities (Alhakami & Slovic, 1994; Finucane et al., 2000; Slovic et al., 2002). However, risk judgments and behavior decisions in the real world most often focus on specific instances of objects and activities that are distinguished by contextual information, details, or constraints (e.g., I like to ski this specific mountain under these particular conditions and consider it safe action) rather than judging their categorical representations (e.g., I like backcountry skiing and consider it a safe activity). Although it is reasonable to assume that affective evaluations and risk perceptions can differ between levels of perception, does the relation between the two found at the categorical level apply to and influence affective evaluations and risk judgments at the contextual level? That is important because when faced with a specific potential hazard and high uncertainty, decision makers should ideally decide behaviour based on a contextual judgment rather than their general, category-level orientation.

Backcountry skiing is an increasingly popular activity in which people voluntarily engage, most often for purely recreational purposes. The very behaviour of going backcountry skiing indicates that backcountry skiers like the activity and consider it safe at the categorical level.

This reflects the positive relation between liking and perceived safety that Alhakami and Slovic (1994), Finucane et al (2000), and Slovic et al. (2002) found when studying the judgments of categorical representations of objects and activities independent of specific contextual information, details, or constraints. That a skier likes to ski in the backcountry and generally considers it a safe category of activity are certainly sentiments that influence their decisions to ski in the mountains. However, once in avalanche terrain, that skier must selectively attend to affective and cognitive information cues to judge the risk in that specific situation. Does the positive relation between liking and perceived safety demonstrated at the category level hold when a skier perceives a specific instance of backcountry skiing distinguished by contextual details? We investigated whether both contextual valence and contextual perceived risk diverge from categorical perspectives in a way that defies or upholds the positive relation between liking and perceived safety.

Paper 2 examines the relation between affective evaluations (i.e., integral affect) and risk perception across multiple judgments of specific instances (i.e., different scenarios of backcountry skiing). We tested whether objectively uncertain instances of backcountry skiing are disliked and perceived as unsafe, in accordance with the liking-perceived safety relation, despite the population liking the activity and perceiving it as safe at the categorical level. The positive relation between liking and perceived safety was found to hold across multiple judgments of a series of hypothetical scenarios depicting specific instances of the activity. This indicates that contextual valence and perceived risk can dynamically diverge from categorical valence and perceived risk when judging specific instances of that category.

2.3 Paper 3: Confidence in Relation to the Perceived Source of Uncertainty

Faced with uncertainty, a decision maker's belief in the accuracy of their decision influences the likelihood that that decision is translated into action (Fitzgerald et al., 2017; Gill et al., 1998; Koriat & Goldsmith, 1996). It is therefore important that a decision maker's confidence in their risk judgments be justified and accurately calibrated. Otherwise overconfidence can lead to an increased likelihood of dangerous behaviour. However, calibrating confidence is extremely challenging during learning and skill development (Sanchez & Dunning, 2018, 2020) and contention persists about the origin of confidence and the internal effects that drive its

(mis)calibration (Boldt et al. 2017; Petrusic & Baranski, 2003; Meyniel et al., 2015; Yeung & Summerfield, 2012). The confidence of experienced decision makers in complex, real-world decision tasks that typify behaviour is relatively unexamined as compared to the multitude of studies involving novice decision makers (i.e., student samples) performing simple tasks with well-defined parameters of information, such as confidence in general knowledge trivia (Koehler et al., 2002; Yeung & Summerfield, 2012). We therefore used the context of decision making in avalanche terrain to investigate how backcountry skiers with domain-specific experience formed confidence in the complex task of judging avalanche risk under high uncertainty.

Research has demonstrated that confidence increases with domain-specific experience (Fischer & Budescu, 2005; Sanchez & Dunning, 2018, 2020). Although experience is often intuitively associated with better judgments, the relationship between experience, confidence, and judgment accuracy is in fact far more complicated. As explained in the earlier section when describing the decision environment, the stochastic nature and inherent unpredictability of avalanche terrain make it virtually impossible to eliminate environmental uncertainty when assessing avalanche risk. Experience with that judgment task and its information environment should relate directly to an increased understanding of that inherent environmental uncertainty: it is impossible to out-think or solve the uncertainty of some conditions. Are more experienced individuals better able to perceive the inherent environmental uncertainty when judging avalanche risk? If so, do they have greater confidence in that perception? Paper 3 explores the role of domain-specific experience in the formation of confidence in risk judgments in relation to the perceived source of uncertainty for that judgment task. Domain-specific experience was found to magnify confidence. However, although greater experience relates to greater confidence in risk judgment, greater experience does not necessarily correspond to more accurate risk judgments or better calibrated confidence.

Beyond this examination of confidence in relation to experience and the perceived source of uncertainty, we also explore the role of affective evaluations in the formation of confidence. Since affective evaluations are understood as influencing the risk judgment process (Blanchette & Richards, 2010; Finucane et al, 2000; Lerner et al., 2015; Lowenstein et al., 2001; Schwarz, 2012; Slovic, 1987; Slovic et al., 2002, 2004), they might similarly play a role in the formation of confidence in that risk judgment. In Paper 2, we found that the more a person likes a scenario, the safer it is perceived, whereas the more a person dislikes a scenario, the more dangerous it is perceived. However, despite this positive relation, it is possible that an individual likes a

stimulus that is perceived as dangerous or dislikes a stimulus perceived as safe. In Paper 3, we explore whether a congruence (e.g., liking a scenario perceived as safe) or an incongruence (e.g., liking a scenario perceived as dangerous) between affective evaluations and risk judgments influences the formation of confidence. Affective evaluations were found to play a role in the formation of confidence. Confidence was higher when the decision maker's affective evaluation matched their risk perception, and less confident when there was a mismatch between the two.

3 Methodological Approach

The presentation of the methods used in this thesis is divided into two parts: (*i*) an explanation of the methodological approach of the thesis (the present section, Section 3) and (*ii*) a comparative overview of the methods and materials used in each of the three papers (Section 4). In the present section I explain our methodological approach across all three papers to conduct research on risk judgments and decisions using the context of decision making in avalanche terrain. This comprises three sub-section. I first explain the framework we used to study applied risk judgments and decisions. I then explain the quasi-representative design approach we used when developing the hypothetical scenarios. Finally I explain why and how we designed uncertain scenarios through the configuration of information cues.

3.1 Framework for Researching Applied Judgment and Decision Making

Studying applied decision making involves three complementary forms of research: 1) examining the properties of the decision task and its environment, 2) studying how decision makers naturally respond to that task, and 3) investigating methods to help decision makers do better (Fischhoff & Broomell, 2020). This perspective is rooted in social judgment theory (Hammond et al., 1975), a meta-theory that serves as a framework to guide research on applied judgment and decision making rather than providing testable hypotheses (Brehmer & Brehmer, 1988; Brehmer & Joyce, 1988; Cooksey, 1996; Dhami et al., 2004; Dhami & Mumpower, 2018; Hammond et al., 1975). According to that framework, research should aim to describe judgment and decision making processes in order to develop interventions to improve performance of that task. Those interventions should seek to empower decision makers to make better decision or to provide needed protections when better decisions are impossible (Fischhoff & Broomell, 2020). The model of the decision environment and the model of how decision makers naturally respond to the decision task are crucial benchmarks for determining how judgements and decisions can be improved (Brehmer & Joyce, 1988; Dhami & Mumpower, 2018). The present thesis is focused on the second of these three complementary forms of research: investigating three aspects of the avalanche risk judgment and decision process in order to better understand how individuals with domain-specific experience naturally make judge risk and decide behaviour under conditions of high uncertainty. In the general discussion further below, I

consider how my research contributes to investigating methods to help decision makers do better, both for backcountry skiers in avalanche terrain and more broadly in other fields involving applied risk judgments and decisions.

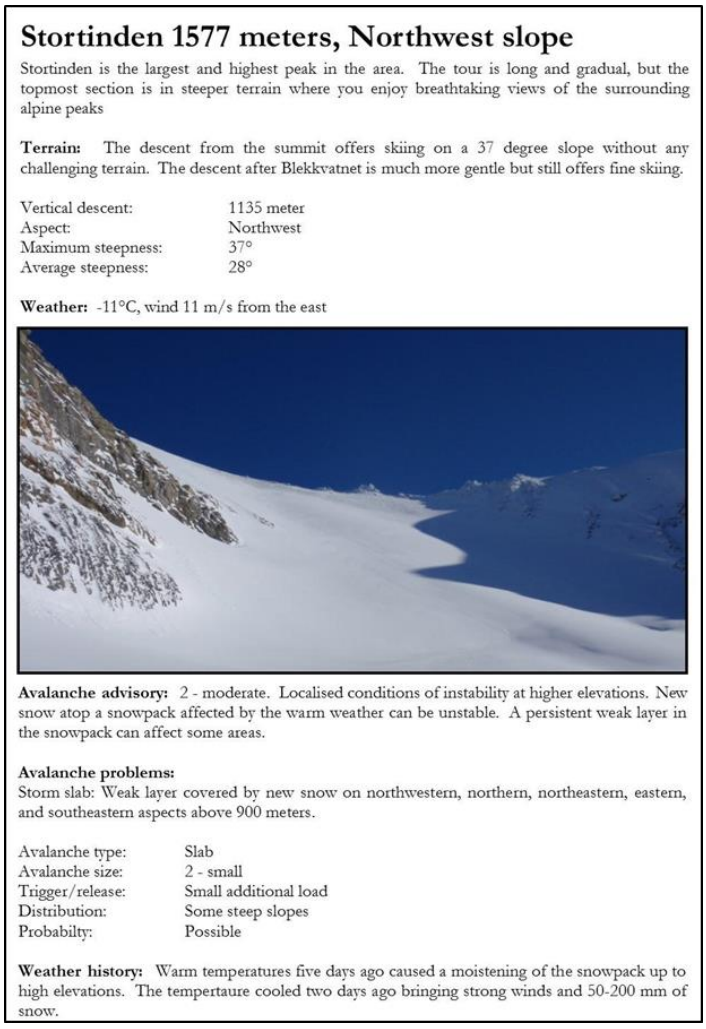
Studying how decision makers naturally respond to an applied decision task is dependent upon a sound understanding of the properties of that task. If we aim to design experiments that have the necessary realism to activate natural risk judgment and decision processes, we must understand the information properties of the real-life decision environment (Cooksey, 1996; Dhami et al., 2004; Hammond, 1966; Hammond et al., 1975). For that purpose, we availed ourselves of the knowledge and expertise within CARE (to which my thesis is connected) and the literature in the field of avalanche safety. Using existing research on the multiplicity of information cues used for risk judgments and decisions in avalanche terrain (Engeset et al., 2018; Haegeli et al., 2010; Landrø et al., 2020a, 2020b; McCammon & Hägeli, 2007; Tremper, 2008), my extensive personal experience in avalanche terrain, and the invaluable support of CARE-affiliated avalanche experts, notably Markus Landrø (M. Landrø, personal communication, June 2017, and September 2019 to January 2020), we sought to design hypothetical scenarios that realistically represent the information properties of avalanche terrain and the behaviour-relevant, goal-oriented judgments and decisions that backcountry skiers make.

3.2 Quasi-representative design

As I previously explained in the introduction (Section 1), the decision environment is central to my thesis. However, due to the ethical concerns with sending research participants into avalanche terrain under conditions of high uncertainty to study their risk judgments and decisions, we used an experimental approach rather than field studies. We utilized a representative design approach of developing realistically complex hypothetical scenarios to ensure high ecological validity in the response of participants for whom the task and environment are familiar and meaningful. Rather than manipulate a single variable in an extremely information-reduced context, which could lead to overestimating the effects of that single factor (for a discussion see Dhami et al., 2004; Koriat, 2018), we manipulated the information cues within those scenarios while maintaining the complexity and richness of information environment in which the decision task is naturally situated.

The studies conducted in Paper 1 used a series of scenarios of one design, while the studies conducted in Papers 2 and 3 used a series of scenarios of another design. We used different scenario designs because of the different contexts for data collection between Paper 1 and Papers 2 and 3. Paper 1 involved a series of studies conducted in an auditorium during seminars on avalanche safety. The time available during the seminars to conduct the studies was limited so we excluded text from the scenarios so that participants could process the available information more quickly. By contrast, the studies in Papers 2 and 3 were conducted online under no circumstantial time constraints. Participants could use as much time as they needed to read the text in each seminar. Figure 1 presents an example of the scenarios used in Paper 1. Figure 2 presents an example of the scenarios used in Papers 2 and 3.

Figure 1. *Example of the scenarios used in Paper 1*
Note: Information cues are communicated through the image and through the icons indicating the slope angle (the first, topmost icon), the forecasted regional avalanche danger level for that scenario (the second icon), and the prevailing regional avalanche problems (the last two icons indicate that there are two potential avalanche problems in the scenario). The icons used in the scenarios are standardized icons defined by the European Avalanche Warning Services (EAWS) and used by the Norwegian Avalanche Warning Service in daily regional avalanche danger forecasts throughout the country.



Stortinden 1577 meters, Northwest slope

Stortinden is the largest and highest peak in the area. The tour is long and gradual, but the topmost section is in steeper terrain where you enjoy breathtaking views of the surrounding alpine peaks

Terrain: The descent from the summit offers skiing on a 37 degree slope without any challenging terrain. The descent after Blekkvatnet is much more gentle but still offers fine skiing.

Vertical descent:	1135 meter
Aspect:	Northwest
Maximum steepness:	37°
Average steepness:	28°

Weather: -11°C, wind 11 m/s from the east

Avalanche advisory: 2 - moderate. Localised conditions of instability at higher elevations. New snow atop a snowpack affected by the warm weather can be unstable. A persistent weak layer in the snowpack can affect some areas.

Avalanche problems:
Storm slab: Weak layer covered by new snow on northwestern, northern, northeastern, eastern, and southeastern aspects above 900 meters.

Avalanche type:	Slab
Avalanche size:	2 - small
Trigger/release:	Small additional load
Distribution:	Some steep slopes
Probabilty:	Possible

Weather history: Warm temperatures five days ago caused a moistening of the snowpack up to high elevations. The tempertaure cooled two days ago bringing strong winds and 50-200 mm of snow.

Figure 2. *Example of the scenarios used in Papers 2 and 3*
Note: Information cues such as the current avalanche danger forecast, slope angle, terrain features, weather history, snowpack characteristics, and the current forecasted hazard(s) are communicated through the text and the image in the scenario.

However, representative design requires that stimuli not only realistically represent the characteristics of the natural decision environment but that they also be a faithful representation of the probabilistic distribution of that environment (Dhami et al, 2004; Dhami & Mumpower, 2018). Fulfilling that design features requires an ecological analysis of the probabilistic structure of the natural decision environment (i.e., the probabilistic distributions of risk and the associated information cues). Unfortunately, such ecological sampling data does not exist for avalanche terrain and the work required to obtain such data is far beyond the scope of my thesis. That the probabilistic distributions of risk and information cues in avalanche terrain varies greatly with annual weather variation and local geographic characteristics makes obtaining such

ecological sampling data all the more challenging. Moreover, the probabilistic structure of the environment also depends upon individual preferences in terms of skiing, risk orientation and acceptance, and group dynamics, to name just a few subjective factors (e.g., Mannberg et al, 2018, 2020). It is therefore difficult to fulfil the design criteria of representative probabilistic distributions prescribed by representative design theory for the decision environment of avalanche terrain.

The studies conducted for Papers 2 and 3 used exclusively uncertain scenarios. That we could not use a sample of scenarios that faithfully represent the probabilistic distribution of risk in the natural decision environment might go some way toward explaining the effects found in those studies, particularly those concerning confidence. People naturally search for patterns among random sequences and are particularly skilled at perceiving patterns but poor at recognizing randomness (Bar-Hillel & Wagenaar, 1991; Oskarsson et al., 2009). It is possible that participants in our studies perceived spurious patterns and correlations across the scenarios they judged with the result that some scenarios appeared safer and others more dangerous. This possibility of perceiving spurious differences in the relative safety/danger between scenarios may have been magnified by the design of the studies themselves. The exclusive use of uncertain scenarios may have been ecologically invalid for a sample of experienced participants who expected a particular distribution of risk among the scenarios that reflected the distribution of risk in their prior real-world experience. Two of the studies conducted for Paper 1 employed safe, uncertain, and dangerous scenarios but, for the reasons explained above, it was impossible to establish whether the proportion of scenarios in those categories represent the probabilistic distribution of the natural environment of avalanche terrain. Recognizing this limitation, I regard my methodology to be based on a quasi-representative design approach.

3.3 Designing Uncertain Scenarios

Being unable to determine and thereby sample a proportional distribution of risk, we purposefully used uncertain scenarios in this research, notably for the studies conducted for Papers 2 and 3. We did so because accidents and fatalities are most frequent during uncertain conditions. Risk judgment and decision errors predominantly occur under conditions of uncertain risk when a decision maker mistakenly judges objectively dangerous conditions as safe. My research is focused on the probabilistic conditions under which accidents are most

likely to occur – under conditions of uncertainty – and therefore designed the scenarios to reflect those conditions.

What makes the conditions in avalanche terrain uncertain and how did we achieve that uncertainty in the hypothetical scenarios? Risk judgments and decisions most often involve making probabilistic inferences based on incomplete and ambiguous evidence (Brehmer & Brehmer, 1988; Brehmer & Joyce, 1988; Brunswik, 1955a; Cooksey, 1996; Dhami et al., 2004; Dhami & Mumpower, 2018; Hammond et al., 1975). Decision makers do not perceive the level of risk in the external environment (i.e., the distal state or stimulus) directly or objectively. Instead, they make an inferential judgment of risk based on multiple proximal cues that imperfectly indicate the true state of that the world (Dhami & Mumpower, 2018; Newel, 2013). As many as 53 information cues might be relevant for risk judgments and decision in avalanche terrain (Landrø et al., 2020a, 2020b). We constructed and presented he scenarios using a selection of those information cues that avalanche experts consider essential for judging avalanche risk. Participants used the proximal information cues presented in the scenarios to judge avalanche risk. As in the real world, they had to interpret macro level (i.e., regional level) information such as the forecasted danger level, weather history, and the prevailing regional avalanche problems and apply that interpretation to evaluate micro level cues such as terrain and slope angle to judge the avalanche risk of a specific slope. None of the scenarios included indicators of either very low or very high risk of avalanche. The available information cues in the scenario photo and the icons (see Figure 1) or textual content and photo (see Figure 2) had limited reliability and validity, and were ambiguous and/or conflicting. The scenario depicted in Figure 2 illustrates how the configuration of information cues create uncertainty. Although the steepness of the slope, the weather history, and the prevailing avalanche problem in the scenario indicate an increased probability of an avalanche, the low danger level of 2-moderate (on a five-category scale), the terrain features, and the lack of any signs in the photograph indicative of the presence of the regional avalanche problem indicate a reduced probability of avalanche. It was not possible to ascertain that the scenario was safe or dangerous. The risk level of the scenario was objectively uncertain given the conflicting evidence. In that way, the available evidence in each of the uncertain scenarios could elicit a range of probabilistic judgments.

4 Methods and Materials

Having explained in the previous section the methodological approach of this thesis, in what follows I provide a comparative overview of the methods and materials used in each of the three papers. I first explain the provenance of the data for each of the three papers. Then I briefly describe the specific methods we used in each paper to investigate the effect of question framing on perceived risk and behaviour intention, the relation between affective evaluations and perceived risk across multiple judgments of specific instances of a category of activity, and the formation of confidence in complex risk judgments. I then present the formulation of the judgment tasks and their response scales as they were presented to participants in each paper. Finally, I present the main model and statistical analysis that we used in each paper.

4.1 The Origin of Data for Papers 1 to 3

Paper 1 is based on empirical evidence collected over six independent, pre-registered studies. Those six studies used the scenario design represented in Figure 1. Data from those six studies are only reported in Paper 1. For Paper 2 and Paper 3, we collected data on risk judgments over a series of five studies using the scenario design represented in Figure 2. As Figure 3 illustrates, Papers 2 and 3 are based on empirical evidence collected over three common studies and one study unique to each paper. The first study in that series of five studies was a pilot study in which we measured risk judgments and confidence but did not measure affective evaluations. As reported in Paper 2, in the next four studies in that series (the second to fifth studies) we measured risk judgments and affective evaluations to test our hypotheses on the relation between those two variables. All data testing the hypothesized relation between affective evaluations and risk judgments are presented in Paper 2. For exploratory purposes, we continued to measure retrospective confidence in all but the fifth and final study in that series. Those data on confidence constituted the basis for Paper 3. Paper 3 explores the formation of confidence using data on risk judgments, affective evaluations, and participant experience from the three common studies reported in Paper 2, and data on risk judgments and participant experience from the first pilot study. All data on confidence and their analysis are uniquely presented in Paper 3, as Table 1 illustrates. Despite drawing upon data from several common studies, the theoretical contextualisation, analyses, results, and their implications are differ between Papers 2 and 3. In Paper 3, we clearly declared that the data on risk judgments,

affective evaluations, and participant experience from three of the four studies were previously published in the context of testing a hypothesized relation between scenario liking and judged safety (i.e., in Paper 2). We further declared that confidence did not pertain to the focus of that investigation and therefore we neither analysed nor published data on confidence in Paper 2. All studies were conducted in accordance with the ethical research protocols of UiT The Arctic University of Norway and the Norwegian Center for Research Data (NSD). Participants gave informed consent prior to all studies.

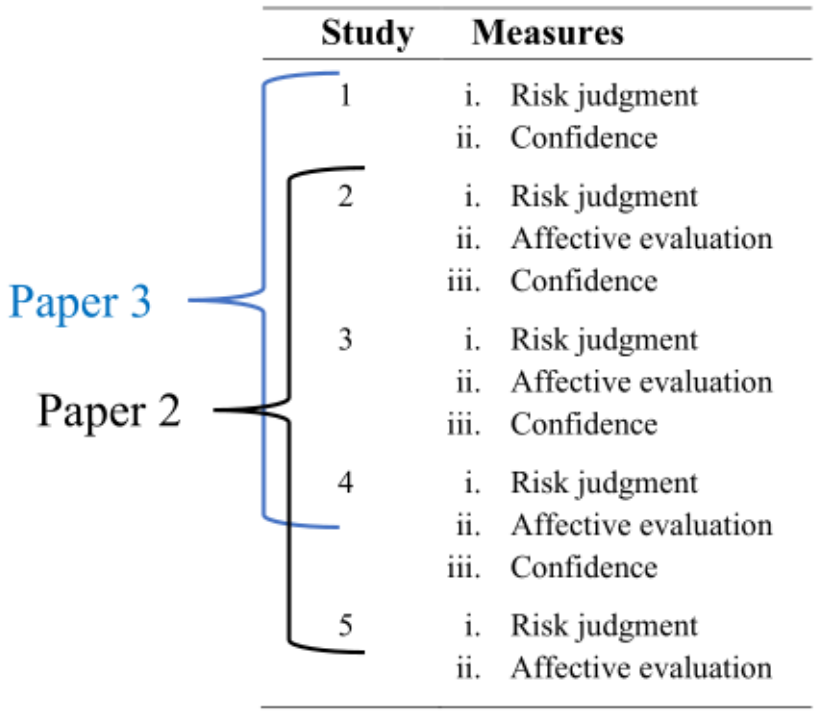


Figure 3. *Provenance of data for Paper 2 and Paper 3*

4.2 Study Designs

Table 1 presents a comparative overview of the studies conducted for each of the three papers. Detailed descriptions of the pre-registrations, participants, study designs, materials, measures, data collection methods, and data preparation and analyses are presented in each paper.

Table 1. *Overview of study designs in Papers 1-3*

Research details	Paper 1: Question framing	Paper 2: Liking & Perceived Safety	Paper 3: Confidence in complex risk judgments
Sample	N = 1599	N = 372	N = 181
Studies	6 studies	4 studies (see Figure 3)	4 studies (see Figure 3)
Stimuli	Studies 1-4: 6 scenarios per participant Studies 5-6: 10 scenarios per participant (see scenario example in Figure 1)	4 scenarios per participant (see scenario example in Figure 2)	Study 1: 6 scenarios per participant Studies 2-4: 4 scenarios per participant (see scenario example in Figure 2)
Design	Multiple stimuli per participant Between participant manipulation: framing of question that elicited the risk judgment	Multiple stimuli per participant Within participant manipulation	Multiple stimuli per participant
Judgment tasks per scenario	1. Risk judgment, framed in terms of safety versus danger 2. Behaviour intention	1. Risk judgments[1] 2. Affective evaluations[1]	1. Risk judgments[1] 2. Affective evaluations[1] 3. Confidence reports[1]
Covariates	Experience score[2] Gender Age	Experience score[2] Gender Age	Experience score[2] Gender Age

Notes: [1] The risk judgments, affective evaluations, and confidence reports involved three measures, the second reversed, calculate a score for each judgment.

[2] Four measures of self-reported skiing ability, avalanche training, years of backcountry skiing, and average number of backcountry skiing days per season were used to calculate a domain-specific experience score per participant. Due to a technical failure during data collection for Paper 1, the covariates were not measured in all studies and are therefore not considered in the analyses and results reported in Paper 1.

4.3 The Judgment Tasks and How They Were Measured

Table 2 presents a comparative overview of the judgment tasks in each paper, the formulation of those tasks as presented to participants, and the response scales or options.

Table 2. *Overview of judgment tasks used in Papers 1-3*

Paper	Judgment task	Formulation of judgment task	Response scale
1	Perceived risk (Between-participant experimental condition)	Safe condition: "How safe is it?"	7-point scale labelled *Not at all safe* and *Completely safe* at the limits
		Danger condition: "How dangerous is it?"	7-point scale labelled *Not at all dangerous* and *Completely dangerous* at the limits
	Behaviour intention	"Would you ski this slope?"	Three response options: *Yes, I cannot say,* or *No*
2 & 3	Perceived risk	"Regarding the avalanche risk…" 1) "Is it safe to ski the slope in these conditions?" 2) "Is it dangerous to ski the slope in these conditions?" (reverse-coded) 3) "Is the snowpack stable enough to ski this slope?"	Three 7-point scales labelled *Not at all* and *Very much* at the limits
	Affective evaluation	"Regarding your desire to ski this slope, the slope appears…" 1) "Attractive" 2) "Uninteresting" (reverse-coded) 3) "Enjoyable"	Three 7-point scales labelled *Not at all* and *Very much* at the limits
	Confidence report	"Regarding your judgments of the avalanche risk…" 1) "Are you confident in your assessment of the avalanche risk?" 2) "Do you doubt your judgment of the avalanche danger?" (reverse-coded) 3) "Are you confident in your judgment of the safety of the snowpack?"	Three 7-point scales labelled *Not at all* and *Very much* at the limits

4.4 Statistical Models and Analyses

Table 3 presents a comparative overview of the main model(s) that we fit and the statistical analysis that we conducted in each paper. Detailed descriptions of all data preparation and analyses are presented in each paper. The data and R script for data processing and analysis for all studies in each paper are publicly available on the Open Science Framework. The relevant links are provided in each paper.

Table 3. *Overview of modelling and statistical analyses for Papers 1-3*

Paper	Models	Statistical analyses
1	Judged safety ~ question frame Behaviour intention ~ judged safety * question frame	Ordinal regression using multilevel models Null hypothesis significance testing Frequentist inferences
2	Judged safety ~ scenario liking	Linear regression using multilevel models Null hypothesis significance testing Frequentist inferences
3	Confidence ~ judged safety * participant experience + scenario liking	Linear regression using multilevel models Parameter estimation using the combined individual participant data from four studies Bayesian inferences

5 Results

In this section, I present qualitative summaries of the results of each of the three papers. The purpose of these summaries is to inform the discussion of those findings and their implication in General Discussion (Section 6). Detailed quantitative reports of all results and discussions of their implications and limitations are presented in each paper.

5.1 Paper 1: Risk Perception and Question Framing

Paper 1 investigated how framing the question that elicits a risk judgment in terms of safety or danger influences the perceived risk of a slope and the intention to ski it. The different theoretical accounts of the framing effect – the valence account, the communication account, and the attention account – make conflicting predictions about the direction and prevalence of an effect evoked by a question frame. Prior research (Payne et al., 2013; Comerford & Robinson, 2017) produced conflicting results about the direction of the question framing effect, and thus the cognitive processes that a question frame is inferred to activate. The findings from this paper make several contributions to advance our theoretical understanding of framing effects. First, this paper establishes the direction of the framing effect when a risk judgment is framed in terms of safety or danger. Across six studies, risk judgments framed in terms of safety (*How safe is it?*) were found to result in more cautious, conservative judgments than when framed in terms of danger (*How dangerous is it?*). The graph on the left in Figure 4 illustrates that framing effect. Second, uncertainty was not required for that effect. There was a question framing effect when judging risk under varying degrees of uncertainty, be it under conditions of objective safety, uncertainty, or danger. These findings suggest that the question frame directed attention in a way that guided selective evidence sampling to evaluate whether the target of judgment meets or fulfils the descriptive state defined by the question frame. The direction of the question framing effect we observed across six studies runs counter to what might be expected based on alternative accounts of framing that assume a valence or communication driven mechanism. Finally, this paper demonstrates the indirect influence of the question frame on behaviour intention. As the graph on the right in Figure 4 illustrates, the probability that participants would ski a slope monotonically increased with an increase in judged safety. Question framing was found to influence judged safety, and judged safety was in turn found to be a powerful determinant of behaviour decisions. By selectively phrasing the

question that elicits a risk judgment, a decision maker's attention can be directed in a way that strategically influences the perception of risk with the effect of making one behaviour outcome more likely.

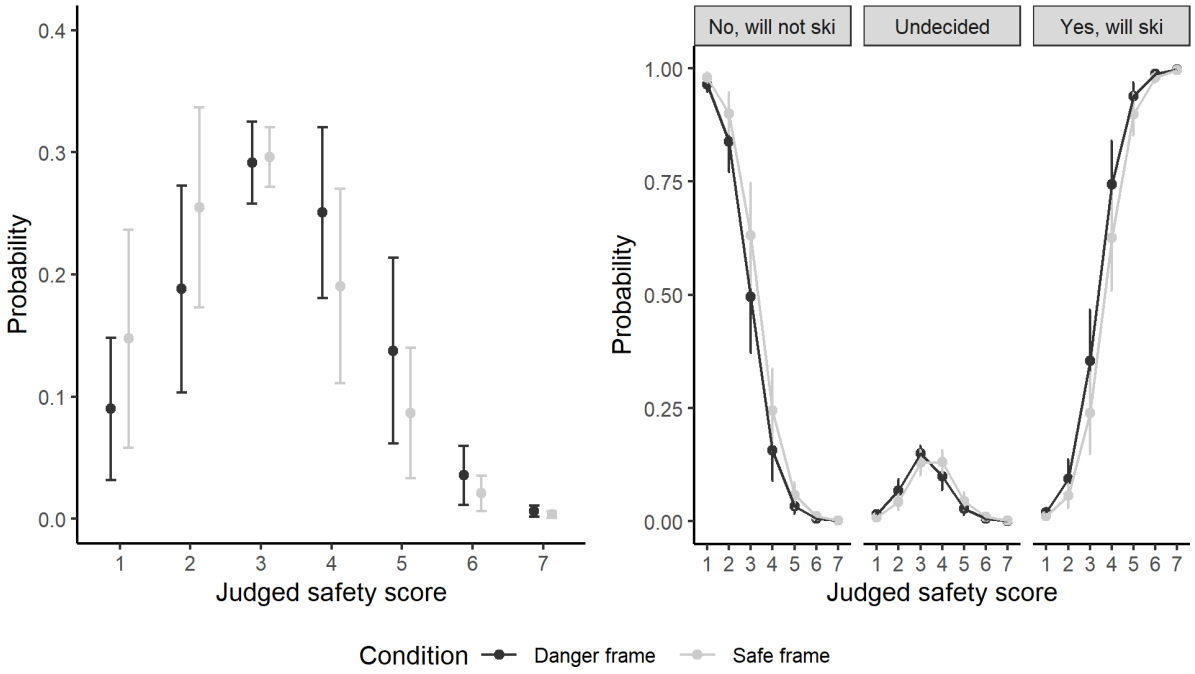


Figure 4. *The effect of question framing on judged safety and behaviour intentions.*
Note: The graph on the left illustrates the probabilities of each judged safety values (1 = *Not at all safe* and 7 = *Completely safe*) by experimental condition. The vertical bars mark the 95% confidence intervals of the probabilities. The graph on the right illustrates the probabilities of each response (faceted into three columns = *No, I will not ski*; *Undecided*; *Yes, I will ski*) per judged safety value and per experimental condition. Each response option is marked at the top of each column in the graph. The vertical bars mark the 95% confidence intervals of the probabilities. Both graphs are plotted using the combined data from studies 1 to 4 in Paper 1 in which only uncertain scenarios were judged.

5.2 Paper 2: Affective Evaluations and Perceived Risk

Paper 2 examined the relation between affective evaluations and risk perception across multiple judgments of specific instances (i.e., different scenarios of backcountry skiing). As Figure 5 illustrates, across four studies we found evidence that higher self-reported liking of backcountry ski scenarios corresponded to judgments of higher avalanche safety (or, conversely, higher disliking corresponded to judgments of higher avalanche danger). Although prior research demonstrated the positive relation between liking and perceived safety when judging

categorical representations of stimuli (Alhakami & Slovic, 1994; Finucane et al., 2000; Slovic et al., 2002), this paper extends those findings by demonstrating that the liking-perceived safety relation holds across multiple judgments of highly uncertain, specific instances of an activity that are distinguished by contextual information. Furthermore, despite the facts that backcountry skiers like backcountry skiing in avalanche terrain and deem the activity to be safe, specific instances of backcountry skiing assumed a negative contextual valence and were perceived as unsafe in accordance with the liking-perceived safety relation. The relation between liking and perceived safety holds at the contextual level of perception, even when the valence and perceived safety of a specific instance of a stimulus conflicts with the established valence and perceived safety of that category of stimulus for the decision maker. Contextual judgments of specific instances of an activity can dynamically diverge from the person's orientation toward that category of activity. However, even though the two levels of judgments can diverge, that the relation between affective evaluations and perceived risk holds at both levels of perception across multiple judgments of specific instances of an activity provides a dynamic picture of the degree to which cognitive and affective factors are intertwined.

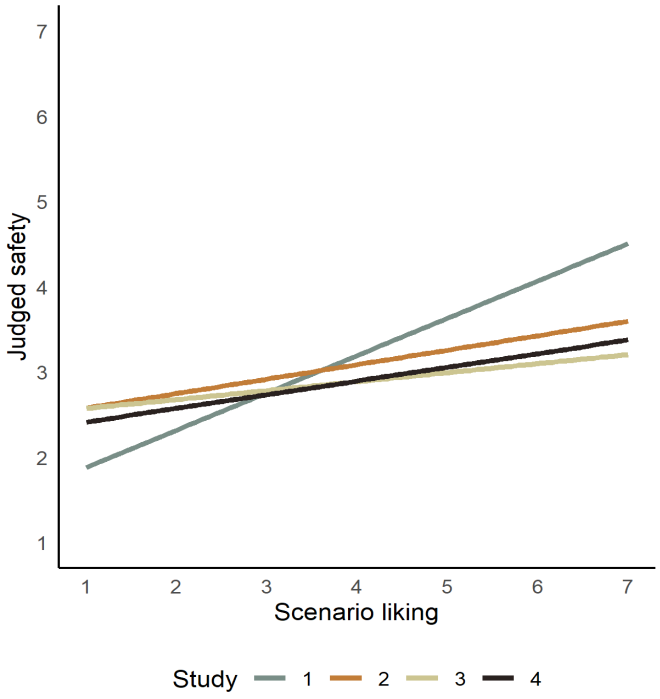


Figure 5. *Scenario liking as a predictor of judged safety per study*

5.3 Paper 3: Confidence in Complex Risk Judgments

Paper 3 explores the roles of the perceived source of uncertainty, domain-specific experience, and affective evaluations in the formation of confidence in complex risk judgments. Domain-specific experience was found to magnify confidence, echoing previous findings on the development of confidence (Fischer & Budescu, 2005; Sanchez & Dunning, 2018, 2020). Yet despite their greater confidence, variance in judged safety did not decrease among more experienced participants. Confidence in judged safety increased with experience despite highly experienced participants demonstrating as much disagreement in judged safety as participants with little experience. Even the most experienced participants did not recognize that the safety or danger of each scenario could not be established. Confidence was lowest at the mid-range of the response scale and highest at either end of the response scale. More experienced participants did not judge with greater confidence that it is impossible to establish the safety or danger of the scenarios. Instead, as Figure 6 illustrates, more experienced participants had greater confidence in their mistaken perceptions that a given scenario was completely safe or completely dangerous. Experience with the judgment task did not increase recognition of the information limitations and inherent uncertainty of the environment.

Affective evaluations were also found to play a role in the formation of confidence. As Figure 7 illustrates, the effect of affective evaluations was specific to whether a scenario was judged as safe to some degree (judged safety ≥ 4) or dangerous to some degree (judged safety < 4). Liking a scenario perceived as safe or disliking a scenario perceived as dangerous increased confidence. By contrast, liking a scenario perceived as dangerous or disliking a scenario perceived as safe reduced confidence in the risk judgment. Participants were more confident when their affective evaluation matched their risk perception, and less confident when there was a mismatch between the two. These findings are evidence of an interplay between affective and cognitive judgments in the formation of confidence under high uncertainty.

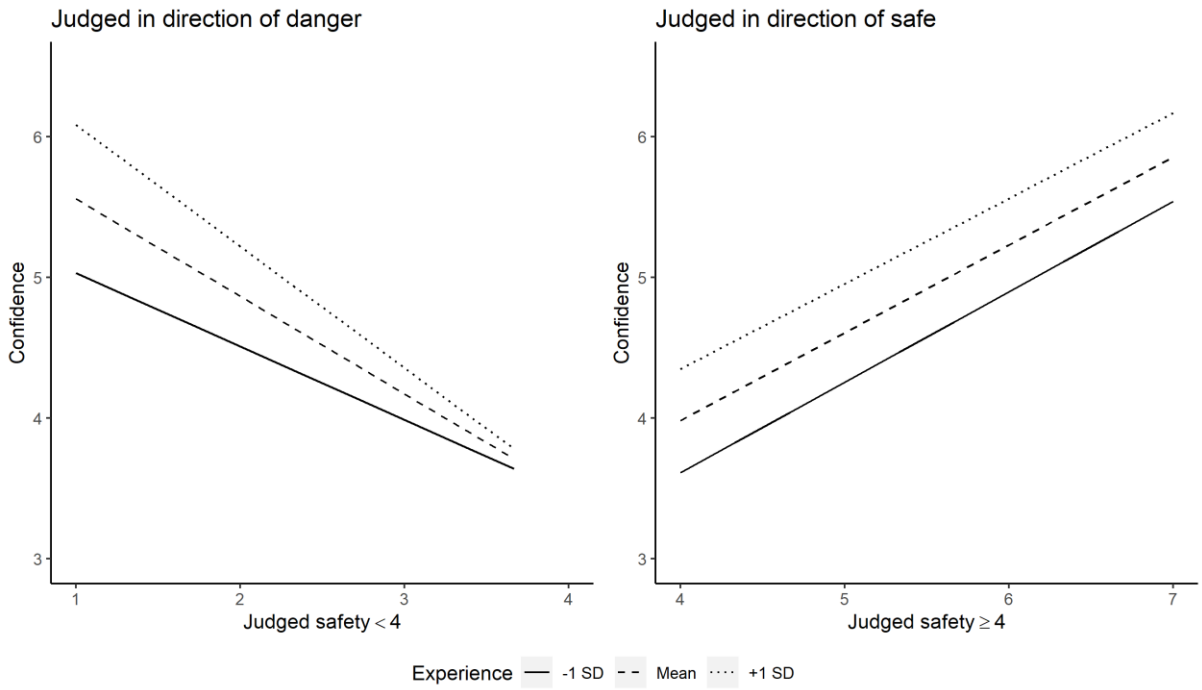


Figure 6. *The relation of confidence to judged safety and experience*

Note: The effects of judged safety and participant experience on confidence (*confidence ~ judged safety * participant experience*) when a scenario is perceived as dangerous to some degree (plot on left, judged safety < 4) or perceived as safe to some degree (plot on right, judged safety ≥ 4). The effect of judged safety is plotted on the x-axis. The effect of participant experience is plotted as a solid line for -1 SD participant experience, a dashed line for mean participant experience, and a dotted line for +1 SD participant experience.

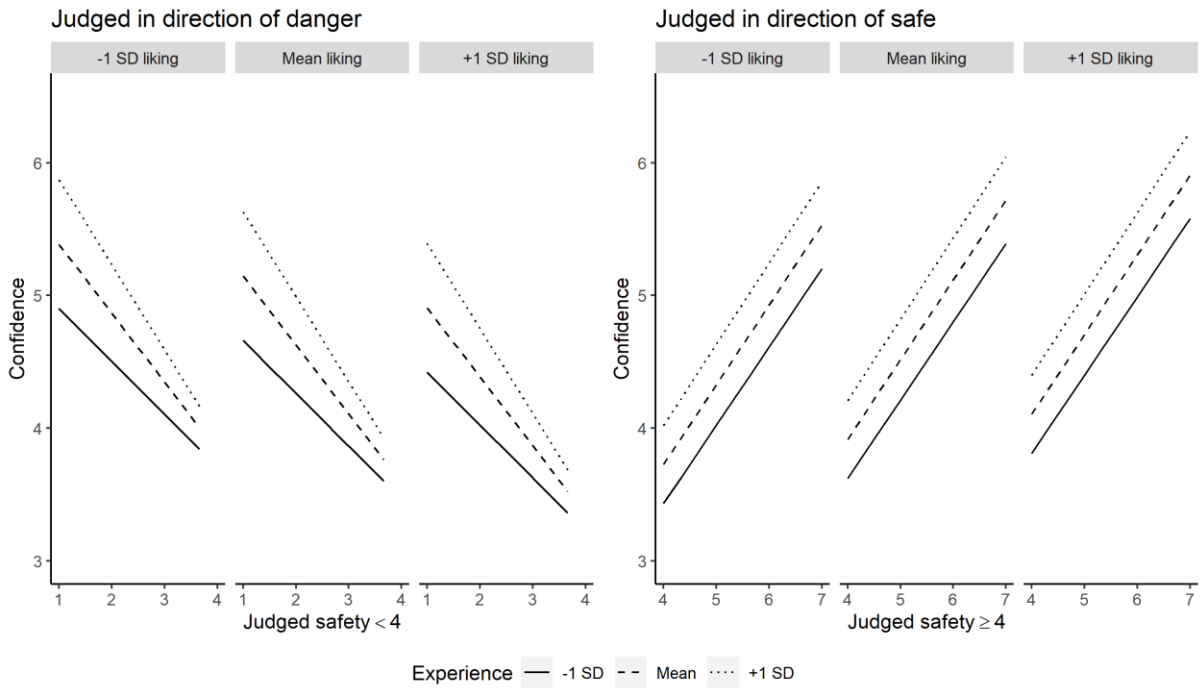


Figure 7. *The relation of confidence to judged safety, experience, and affective evaluations*
Note: The effects of judged safety, scenario liking, and participant experience on confidence (confidence ~ judged safety * participant experience + scenario liking) when a scenario is perceived as dangerous to some degree (plot on left, judged safety < 4) or perceived as safe to some degree (plot on right, judged safety ≥ 4). The effect of judged safety is plotted on the x-axis. The effect of participant experience is plotted as a solid line for -1 SD participant experience, a dashed line for mean participant experience, and a dotted line for +1 SD participant experience. The effect of scenario liking is faceted into -1 SD scenario liking in the left column of each plot, mean scenario liking in the middle column of each plot, and +1 SD scenario liking in the right column of each plot.

6 General Discussion

In this thesis, I have studied three distinct but complementary facets of the risk judgment and decision process using a specific applied context. In three empirical papers, we investigated the effect of question framing on risk perception and behaviour intention, examined the relation between affective evaluations and risk judgments, and explored the formation of confidence in complex risk judgments. The results and implications of those three lines of study are discussed in each of the three papers. I will therefore refrain from unnecessarily repeating those discussions here. Needless to say, I believe that each of the three research papers succeeds in advancing our understanding of a select aspect of the decision making process.

My ambition for this section is to raise the discussion up a level beyond the specific research question(s) addressed in each paper in order to consider the findings from my research in the broader context of applied decision making. Accordingly, I will first consider how the combined findings from the three papers further reveal the crucial role of affective evaluations in multiple stages of the decision making process. I focus on affective evaluations here because these implications only become apparent when considering the combined findings from all three papers, and so is not discussed in any one paper. I also discuss the limits of those findings and propose directions for future research on integral affect. I then consider the relevance of the findings from each of the three papers for interventions to help decision makers do better, both in avalanche terrain and more broadly. I begin by first discussing how our findings might inform interventions to help people make more accurate risk judgments, but also consider the challenges and limitations of that pursuit in a context such as decision making in avalanche terrain. I then discuss the relevance of our findings in relation to an alternative conceptualisation of "better" decisions based on a consequentialist interpretation of decision quality. I consider how our research contributes to promoting the safest decision given the available evidence rather than the most objectively accurate decision.

6.1 The importance of affective evaluations

Traditional theories of rational choice eschewed the evaluative and informative role of feelings in the decision making process, regarding them as a mere consequence of logical cognitive processing (Lerner et al, 2015). Yet theoretical and empirical work in psychology and

economics now suggest that affect is a key mechanism that informs, guides, or influences risk judgments (Blanchette & Richards, 2010; Damasio, 1994; Lerner et al, 2015; Lowenstein et al., 2001; Schwarz, 2012; Zajonc, 1980). The findings from my research provide further evidence of the prominent role of affective evaluations in the judgment and decision making process. Appraising the combined results of each of my three papers, it is possible to make inferences about the process by which affective evaluations indirectly influence behaviour.

Behaviour is what ultimately exposes people to risk. Affective evaluations play a crucial role in many stages of the decision process that culminates in behaviour. In Paper 2, we found that risk judgments relate to affective evaluations: the more a person liked a specific highly uncertain context of backcountry skiing in avalanche terrain, the safer they judged it to be. Although my research failed to demonstrate that liking had a causal influence on perceived safety, other literature provides empirical evidence and theoretical justification for such a causal relation (Blanchette & Richards, 2010; Damasio, 1994; Finucane et al., 2000; Slovic et al., 2002; Lerner et al., 2015; Lowenstein et al., 2001; Schwarz, 2012; Slovic et al., 2004; Zajonc, 1980). In Paper 1, we found that risk judgments are a powerful determinant of behaviour intention: the likelihood that an individual would choose to ski a slope increased monotonically with judged safety. When considered together, the separate findings from Papers 1 and 2 suggest that affective evaluation may indirectly influence behaviour. Based on (*i*) the evidence from Paper 2 for the relation between the affective evaluation and the perceive risk of a specific, contextually situated stimulus, (*ii*) prior empirical evidence and theoretical justification that affective evaluations influence risk judgments, and (*iii*) the evidence from Paper 1 that risk judgments directly determine behaviour intention, we can infer that affective evaluations indirectly influence behaviour via their direct influence on perceived risk. Future research that directly tests that inferred indirect causal relation would make a valuable contribution to better understanding the role of affective evaluations in the decision making process.

Another process by which affective evaluations appear to influence behaviour, and ultimately risk exposure, is via the formation of confidence in a risk judgment. In Paper 3, we found that affective evaluations influence confidence. The more a decision maker likes a scenario perceived as safe or dislikes a scenario perceived as dangerous, the greater their confidence in that risk judgment. It is the first of those two formulations that is presently relevant since that is the behaviour that results in risk exposure. Confidence in a decision is the control process for whether that decision is translated into action (Fitzgerald et al., 2017; Gill et al., 1998; Koriat & Goldsmith, 1996). The likelihood that a skier skis a slope increases with their confidence in

their perception of the risk of the slope. Not only do affective evaluations indirectly influence behaviour through the relation with perceived risk, affective evaluations also indirectly influence behaviour through a direct influence on confidence. The results across all three papers provide a dynamic picture of the degree to which cognitive and affective evaluations are intertwined in the process of judging risk and ultimately deciding behaviour.

A clear limitation of this thesis is that we were unable to experimentally demonstrate the causal relation between affective evaluations and risk judgments and to elucidate how affective evaluations emerge in response to a stimulus. In the studies conducted for Paper 2, we attempted to manipulate the attractiveness of the backcountry skiing scenarios while holding the risk level constant with two different experimental designs: manipulating the presence versus absence of ski tracks and manipulating the sunny versus overcast sky in the scenario photo. The purpose of those manipulations was to investigate if the subjective affective value of a single feature of the scenarios influences the degree of scenario liking and, consequently, judged safety. Unfortunately, both manipulations failed to influence scenario liking, making it impossible to test causality. In Paper 2, we discuss some interesting consequences of those failed manipulations.

How affect emerges in response to a stimulus is an important consideration when studying the relation between affective evaluations and risk judgments. Affective responses can be evoked by the real or imagined features of a judged stimulus (Cohen et al., 2008). Each of those features can have a distinct subjective affective value that influences—to a greater or lesser degree—the global liking evaluation of the target stimulus. In that way, liking, as a global affective response, is the aggregation of the affective values of a stimulus' multiple features. This aligns with the theoretical perspective in the field of attitude research that global attitudinal responses (liking-disliking) are generalized evaluative judgments based on an integration of multiple features of the target stimulus and the context in which it is perceived (Schwarz, 2007, 2012). The affective value of some things may be virtually universal such as the negative response to rotting food, whereas the varied affective responses evoked by other things may be far more subjective. Additionally, some feature(s) of a judged stimulus may elicit greater affective responses according to the ease with which they are ascribed affective value, making them more easily used within the judgment process (Wilson & Arvai, 2006). Finally, some feature(s) of a stimulus might also have a "stigmatizing" association by which extreme weight or importance, either positive or negative, is ascribed to a single feature that influences the integration of all other affective features of the stimulus (Finucane & Holup, 2006).

It has been proposed that the relation between affective evaluations and risk judgments is a beneficial cognitive mechanism that facilitates the processing of complex evidence to enable quick and efficient judgments (Clore & Huntsinger, 2007; King & Slovic, 2014; Slovic et al., 2005). Attending to an affective response should increase the normative accuracy of risk judgments when that affective response is elicited by the content of the judgment task (Blanchette & Richards, 2010; Pham, 1998; Schwarz, 2012). Yet research on the role of product appearance in consumer choices found that the symbolic value of a single feature of a product can influence product evaluation and consumer choice, despite that feature having no bearing on product functionality (Creusen & Schoormans, 2005). Does the influence of the affective value of a single feature of a stimulus extend beyond consumer evaluations to judgments of risk and personal safety? Might the affective value of a single judgment-irrelevant feature of a complex, highly uncertain, and potentially deadly situation bias the perceived safety of that situation? Answers to these questions are needed in order to better understand the role of affective evaluations in risk judgments. It is crucial to ascertain whether the affective evaluation influencing a risk judgment conveys complex information borne of valid experience about the value or quality of the stimulus, or if in some circumstances an affective evaluation can instead be the mere expression of motivational aspects that are irrelevant to the risk judgment.

6.2 Better risk judgments and decisions

Uncertainty and risk are ubiquitous characteristics of life, and the ability to adequately judge and react to risk is essential for a person's survival and success. Fortunately judgment and decision making researchers are here to help. As previously mentioned, studying applied decision making involves three complementary forms of research: 1) examining the properties of the decision task and context, 2) studying how decision makers naturally respond to that task, and 3) investigating methods to help decision makers do better (Fischhoff & Broomell, 2020; Hammond et al., 1975). Although the present thesis focused on the second form of research – increasing our understanding of three aspects of the risk judgment and decision process under conditions of high uncertainty – the findings have implications for future research that aims to identify and test interventions to help decision makers do better.

The quality of a decision maker's performance can be evaluated in terms of its accuracy or its coherence. Decision accuracy concerns the degree to which the decision maker's perception

corresponds to the true state of the world (i.e., correspondence theory, see Brunswik, 1955b). Decision coherence concerns reasoning in accordance with the rules of logic and probability theory in a consistent, coherent manner (Fischhoff & Broomell, 2020). For tasks like decision making in avalanche terrain that involve inferring the likelihood of negative outcomes and then deciding the most appropriate behaviour, the degree to which the decision maker's perception corresponds with environmental reality is the most applicable measure of performance quality. Interventions should seek to empower decision makers to avoid (or at least reduce the frequency and/or likelihood) mistakenly perceiving objectively dangerous conditions as safe.

A focus on failed decision processes resulting in avalanche accidents has dominated the study of decision making in avalanche terrain for the past two decades (Johnson et al., 2020). By examining how backcountry skiers make inaccurate judgements, researchers have tried to understand the causes for judgement errors. Efforts to help backcountry skiers make better decisions in avalanche terrain have focused on improving the objective accuracy (i.e., the degree of correspondence with environmental reality) of risk judgments (Johnson et al., 2020; Landrø et al. 2020a, 2020b). National avalanche warning services provide detailed avalanche forecasts and information resources to increase and improve publicly available knowledge about the potential dangers in avalanche terrain (Engeset et al., 2018). Researchers and avalanche experts have developed (and continue to develop) a variety of decision making frameworks to support risk judgments and decisions (Landrø et al., 2020a, 2020b; McCammon & Hägeli, 2007). Avalanche safety educators train people, often according to established national standards, in the use of those information resources and decision making frameworks. Interventions targeting affective evaluations might have a role in those efforts to improve decision accuracy.

In Paper 2, we found that affective evaluations relate to perceived risk across multiple, context-specific judgments (a relation that, as I previously explained, we might regard as causal). Affective evaluation can be more or less accurate with respect to the objective risk. Interventions that improve the accuracy of affective evaluations should correspondingly increase the normative accuracy of risk judgment. How can that be achieved? Affective evaluations can boost decision accuracy if the affective response is evoked by relevant, valid content of the judgment task (Blanchette & Richards, 2010; Pham, 1998; Schwarz, 2012). The accuracy with which an affective evaluation focuses on valid, relevant information cues should, like any other experiential learning, be developed and calibrated through valid experience involving informative, timely, and representative feedback (Hogarth et al., 2015). Further

research is required to better understand this process of calibrating of affective evaluations through experiential learning. Interventions that improve the validity and representativeness of feedback during the learning and skills refinement process should improve the calibration of affective evaluations and, as a result, the accuracy of risk judgments. Furthermore, as I explained above, research is needed to investigate if the affective value of some features of the stimuli bias and distort affective evaluations and, consequently, risk judgments. If affective evaluations are found to negatively bias risk judgments (e.g., an overly positive affective evaluation causes an unsafe situation to be judged safe), interventions should aim to improve their calibration and/or reduce their role within avalanche risk judgments. Should the latter strategy be pursued, researchers will face the challenge of mitigating the effect of affect within the decision process without unnecessarily diminishing people's enjoyment of the activity or object of judgment.

6.3 An alternative approach to better risk judgments and decisions

It is difficult to study avalanche risk judgment performance in terms of accuracy because it is virtually impossible to establish the base rate *p*(snow instability) and the false alarm rate *p*(snow instability | avalanche not triggered), which are necessary for determining accuracy. Nonetheless, accident and fatality rates indicate that inaccurate risk judgments persist (Birkeland et al., 2017; Jekich et al., 2016; Niedermeier, et al., 2020). The various efforts to improve the accuracy of avalanche risk judgments mentioned above have failed to establish an avalanche prediction method that is free from false-positive or false-negative results (Simenhois & Birkeland, 2009; Landrø et al., 2020a). Reducing avalanche accidents and fatalities continues to elude educators and researchers (Furman et al., 2010) and it would be naïve to expect new interventions targeting affective evaluations (see above) to be a panacea that finally solves the problem of mistaken risk judgments. Might pursuing interventions to improve accuracy – understood as the degree to which the decision maker's perception corresponds with environmental reality – be a frustratingly difficult goal that henceforth can furnish only diminishing returns? Might the uncertainty, complexity, and variability of the decision environment mean that a satisfactory level of accuracy is unachievable for most people?

An alternative strategy would be to adopt a broader conceptualization of the benchmark for decision accuracy. A consequentialist interpretation of rationality in cognition measures accuracy in terms of success in achieving one's goal in the world (Kozyreva & Hertwig, 2021; Schurz & Hertwig, 2019). In avalanche terrain, this goal would be *to not trigger an avalanche* rather than to correctly ascertain the objective degree of avalanche risk. This measure of accuracy concerns the degree to which a decision corresponds with reality *to the extent that* the decision maker achieves their goal under the constraints and affordances of the decision environment and the capacity of the decision maker (Kozyreva & Hertwig, 2021). Interventions based on a consequentialist interpretation of accuracy should seek to empower decision makers to avoid (or at least reduce the frequency and/or likelihood) the negative potential outcome.

Better decisions in avalanche terrain can be understood in terms of safety rather than their correspondence with the true state of the world. The ultimate aim of the risk judgment and decision process in avalanche terrain is to minimize the likelihood of making a "*go*" decision under objectively "*no-go*" conditions. Better decision making can therefore be achieved by increasing the likelihood of a specific behaviour, namely the decision to not ski when internal and/or external uncertainty is high. Such an outcome measure of decision quality is common in the fields of marketing (Biswas, 2009; Biswas & Grau, 2008; Levin & Gaeth, 1988) and communication (Block & Keller, 1995; Chong & Druckman, 2007; Scheufele & Tewksbury, 2007) where promoting a specific decision outcome defines the criteria for measuring quality or success of the intervention. Interventions based on a consequentialist interpretation of decision quality would aim to ensure the safest decision given the available evidence rather than the most objectively accurate decision given the evidence.

The discussion section in Paper 1 considers how question framing can be harnessed to boost people's natural decision making competencies in order to ensure safer risk perceptions and behaviour. Those findings have the potential to inform the development of policies and practices that harness question framing in domains of applied risk perception and risk communication. The applied relevance of question framing is discussed in detail in Paper 1. Interventions targeting confidence represent another possibility to promote more cautious and conservative "go" decisions in the face of high uncertainty. Confidence, a decision maker's belief in the accuracy of their decision, influences the likelihood that that decision is translated into action (Fitzgerald et al., 2017; Gill et al., 1998; Koriat & Goldsmith, 1996). Although improving the accuracy of risk judgments and decisions in avalanche terrain is extremely challenging, interventions to improve the calibration of confidence and mitigate overconfidence

represent an opportunity to reduce the likelihood that a mistaken risk judgment is translated into dangerous behaviour. By understanding the role of various internal signals in the formation of confidence, we might identify processes that we can leverage to reduce (over)confidence and thereby increase the likelihood of safer behaviour. For example, the results from Paper 3 suggest that decision makers mistakenly perceive experience as corresponding to expertise. Interventions that make people more aware of the limitations of their experience and the frequent invalid feedback from the environment might contribute to diminishing overconfidence and reducing the likelihood that incorrect risk judgments are translated into action. A second possible intervention might target the relation between affective evaluations and confidence. Paper 3 found that the more a participant liked a scenario, the more confident they were in their judgment that that scenario was safe. If further research on the accuracy of affective evaluations (see above) found that liking reflected motivational factors rather than a risk-relevant evaluation, then interventions to reduce its dangerous magnification of confidence would be appropriate.

This strategy of promoting the safest decision rather than the most accurate decision might resemble a perspective of paternalistic liberalism (Thaler & Sunstein, 2008). It assumes that one risk perspective and behaviour decision is better, irrespective of accuracy (i.e., degree of correspondence with external reality), and emphasizes increasing the likelihood of that behaviour. Such a paternalistic liberal perspective seems expedient given the complexity of decision making in avalanche terrain, the uncertainty of the decision environment, and the failures of prior interventions to ensure risk judgments and decisions of sufficient objective accuracy. Importantly, however, my proposal to establish safety as the measure of decision quality is not based on the perspective that human cognition is inherently flawed. Nor does it constitute a nudge that systematically manipulates the decision environment (i.e., choice architecture) to influence behaviour in a specific manner (Fischhoff & Broomell, 2020; Hertwig et al., 2019; Thaler & Sunstein, 2008). Rather, through transparent interventions that make small, incremental advances, this approach aims to promote better, safer decision making processes from which better, safer behaviour follows. By defining safety rather than accuracy as the measure of decision quality, the present research aims to boost people's natural decision making competencies (Hertwig et al., 2019) to promote the safest decisions given the available evidence.

Fortunately, promoting the safest decision does not come at the cost of decision accuracy. In the case of backcountry skiing in avalanche terrain, increased accuracy and increased likelihood

of performing a specific behaviour are highly correlated. As the accuracy of avalanche risk judgments increase, the likelihood of making a safe decision about whether to ski also increases. We need not sacrifice accuracy in pursuit of increased likelihood of safer behaviour. By understanding the myriad cognitive processes that are activated during avalanche risk judgments and decisions, it is possible to design interventions that decrease the frequency of accidents and deaths by 1) increasing accuracy, 2) increasing the likelihood of more cautious behaviour, or 3) a combination of those two. The community of avalanche safety professionals has already taken note. At their invitation, the findings on question framing from Paper 1 were published as a popular science article in the spring 2021 volume of *The Avalanche Journal* (Stephensen et al., 2021). That same article will be re-published in a forthcoming volume (June 2021) of the *New Zealand Avalanche Dispatch*. See the appendix for the text of that article. This strategic approach to promoting better decisions has the potential for wide implementation in a number of applied disciplines. By understanding the factors in the decision process that predict whether a certain behavioural outcome is more or less likely, we can design interventions that harness, exploit, or leverage specific cognitive processes that are active during the decision process to increase the likelihood of a desired outcome.

7 Concluding Remarks

In this thesis, I used a context of applied decision making as the framework for designing experiments to study select cognitive processes during the risk judgment and decision process. Decision making in avalanche terrain provided a valuable context to investigate how a population of highly engaged individuals with domain-specific experience perform complex, goal-oriented decision tasks under conditions of high uncertainty. The findings from this research contribute to advancing our understanding of how people make complex risk judgments and decisions under conditions of high uncertainty. Those results also have potential relevance for the development of interventions to help decision makers do better, whether better be understood in terms of risk judgments that more accurately correspond with environmental reality or in terms of behaviour decisions that increase the likelihood of the safest decisions given the available information. The findings from Paper 1 advance our theoretical understanding of framing effects and the cognitive process that a question frame activates. Those findings have the potential to inform the development of practices that harness question framing for applied risk perception and communication. The findings from Paper 2 provide further evidence of the relation between affective evaluations and risk judgments. Those findings provide a dynamic picture of the degree to which cognitive and affective mechanisms are intrinsically intertwined in the decision making process. Finally, the findings from Paper 3 shed light on the roles of uncertainty, domain-specific experience, and affective evaluations in the formation of confidence in complex risk judgments. Those findings shed light on the various internal signals that influence metacognitive judgments.

References

Alhakami, A. S., & Slovic, P. (1994). A Psychological Study of the Inverse Relationship Between Perceived Risk and Perceived Benefit. *Risk Analysis, 14*(6), 1085-1096. https://doi.org/10.1111/j.1539-6924.1994.tb00080.x

Bar-Hillel, M., & Wagenaar, W. A. (1991). The perception of randomness. Advances in Applied Mathematics, 12(4), 428–454. https://doi.org/10.1016/0196-8858(91)90029-I

Birkeland, K. W., Greene, E. M., & Logan, S. (2017). In response to avalanche fatalities in the United States by Jekich et al. *Wilderness & Environmental Medicine, 28*(4), 380–382. https://doi.org/10.1016/j.wem.2017.06.009

Biswas, D. (2009), The effects of option framing on consumer choices: Making decisions in rational versus experiential processing modes. *Journal of Consumer Behaviour, 8*(5), 284-299. https://doi.org/10.1002/cb.288

Biswas, D. and Grau, S.L. (2008), Consumer choices under product option framing: Loss aversion principles or sensitivity to price differentials? *Psychology & Marketing, 25*, 399-415. https://doi.org/10.1002/mar.20217

Blanchette, I., & Richards, A. (2010). The influence of affect on higher level cognition: A review of research on interpretation, judgement, decision making and reasoning. *Cognition and Emotion, 24*(4), 561-595. https://doi.org/10.1080/02699930903132496

Block, L., & Keller, P. (1995). When to Accentuate the Negative: The Effects of Perceived Efficacy and Message Framing on Intentions to Perform a Health-Related Behavior. J*ournal of Marketing Research, 32*(2), 192-203. https://doi.org/10.2307/3152047

Boldt, A., de Gardelle, V., & Yeung, N. (2017). The impact of evidence reliability on sensitivity and bias in decision confidence. *Journal of Experimental Psychology: Human Perception and Performance, 43*(8), 1520-1531. https://doi.org/10.1037/xhp0000404

Brehmer, A., & Brehmer, B. (1988). What have we learned about human judgment from thirty years of policy capturing. In B. Brehmer & C. R. B. Joyce (Eds.), *Human Judgment: The SJT vVew* (pp. 75–114). New York: North-Holland

Brehmer, B., & Joyce, C. R. B. (1988). Human judgment: The SJT view. New York: North-Holland.

Brunswik, E. (1944). Distal focussing of perception: Size-constancy in a representative sample of situations. *Psychological Monographs, 56,* 1–28. http://dx.doi.org/10.1037/h0093505

Brunswik, E. (1955a). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62,* 193–217. http://dx.doi.org/10.1037/h0047470

Brunswik, E. (1955b). In defense of probabilistic functionalism: A reply. *Psychological Review, 62,* 236 –242. http://dx.doi.org/10.1037/h0040198

Chong, D., & Druckman, J. N. (2007). Framing Theory. *Annual Review of Political Science, 10*(1), 103-126. doi:10.1146/annurev.polisci.10.072805.103054

Clore, G. L., & Huntsinger, J. R. (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences, 11*(9), 393-399. https://doi.org/10.1016/j.tics.2007.08.005

Cohen, J. B., Pham, M. T., & Andrade, E. B. (2008). The nature and role of affect in consumer behavior. In *Handbook of consumer psychology*. (pp. 297-348). New York, NY: Taylor & Francis Group/Lawrence Erlbaum Associates.

Comerford, D. A., & Robinson, J. (2017) Die-by Framing both Lengthens and Shortens Life: Further Evidence on Constructed Beliefs in Life Expectancy. *Journal of Behavioral Decision Making*, 30: 1104– 1112. https://doi.org/10.1002/bdm.2027

Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and applications*. San Diego, CA: Academic Press.

Creusen, M. E. H., & Schoormans, J. P. L. (2005). The Different Roles of Product Appearance in Consumer Choice. *Journal of Product Innovation Management, 22*(1), 63-81. https://doi.org/10.1111/j.0737-6782.2005.00103.x

Damasio, A.R. (1994). *Descartes' Error: Emotion, Reason and the Human Brain*. New York: Avon Books.

Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The Role of Representative Design in an Ecological Approach to Cognition. *Psychological Bulletin, 130*(6), 959-988. https://doi.org/10.1037/0033-2909.130.6.959

Dhami, M. K., & Mumpower, J. L. (2018). Kenneth R. Hammond's contributions to the study of judgment and decision making. *Judgment and Decision Making, 13*(1), 1–22.

Ebert, P. A. (2019). Bayesian reasoning in avalanche terrain: a theoretical investigation. *Journal of Adventure Education and Outdoor Learning, 19*(1), 84-95. https://doi.org/10.1080/14729679.2018.1508356

Engeset, R. V., Pfuhl, G., Landrø, M., Mannberg, A., & Hetland, A. (2018). Communicating public avalanche warnings – what works? *Natural Hazards and Earth System Sciences, 18*(9), 2537-2559. https://doi.org/10.5194/nhess-18-2537-2018

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making, 13*(1), 1-17. https://doi.org/10.1002/(SICI)1099-0771(200001/03)13:1%3C1::AID-BDM333%3E3.0.CO;2-S

Finucane, M. L., & Holup, J. L. (2006). Risk as Value: Combining Affect and Analysis in Risk Judgments. *Journal of Risk Research, 9*(2), 141-164. https://doi.org/10.1080/13669870500166930

Fischer, I., & Budescu, D. V. (2005). When do those who know more also know more about how much they know? The development of confidence and performance in categorical decision tasks. *Organizational Behavior and Human Decision Processes, 98*(1), 39-53. https://doi.org/10.1016/j.obhdp.2005.04.003

Fischhoff, B., & Broomell, S. B. (2020). Judgment and Decision Making. *Annual Review of Psychology, 71*(1), 331-355. https://doi.org/10.1146/annurev-psych-010419-050747

Fitzgerald, L. M., Arvaneh, M., & Dockree, P. M. (2017). Domain-specific and domain-general processes underlying metacognitive judgments. *Consciousness and Cognition, 49*(Supplement C), 264-277. https://doi.org/10.1016/j.concog.2017.01.011

Fox, C. R., & Ülkümen, G. (2011). Distinguishing Two Dimensions of Uncertainty. In W. Brun, G. Keren, G. Kirkebøen, & H. Montgomery (Eds.), *Perspectives on Thinking, Judging, and Decision Making: A Tribute to Karl Halvor Teigen*. Oslo: Universitetsforlaget.

Furman, N., Shooter, W., & Schumann, S. (2010). The Roles of Heuristics, Avalanche Forecast, and Risk Propensity in the Decision Making of Backcountry Skiers. *Leisure Sciences, 32*(5), 453-469. https://doi.org/10.1080/01490400.2010.510967

Garcia-Retamero, R., & Cokely, E. T. (2011). Effective communication of risks to young adults: Using message framing and visual aids to increase condom use and STD screening. *Journal of Experimental Psychology: Applied, 17*(3), 270–287. https://doi.org/10.1037/a0023677

Garcia-Retamero, R., & Galesic, M. (2010). Who profits from visual aids: overcoming challenges in people's understanding of risks [corrected]. *Social Science & Medicine, 70*(7), 1019-1025. https://doi.org/10.1016/j.socscimed.2009.11.031

Gill, M. J., Swann Jr, W. B., & Silvera, D. H. (1998). On the genesis of confidence. *Journal of Personality and Social Psychology, 75*(5), 1101-1114. https://doi.org/10.1037/0022-3514.75.5.1101

Gillies, D. (2000). *Philosophical theories of probability*. London: Routledge.

Haegeli, P., Haider, W., Longland, M., & Beardmore, B. (2010). Amateur decision-making in avalanche terrain with and without a decision aid: A stated choice survey. *Natural Hazards, 52*(1), 185-209. https://doi.org/10.1007/s11069-009-9365-4

Hammond, K. R. (1966). Probabilistic functionalism: Egon Brunswik's integration of the history, theory, and method of psychology. In K. R. Hammond (Ed.), The psychology of Egon Brunswik (pp. 15–80). New York: Holt, Rinehart and Winston.

Hammond, K. R., Stewart, T. R., Brehmer, B. , & Steinmann, D. O. (1975). Social judgment theory. In M. F. Kaplan & S. Schwartz (Eds.), Human judgment and decision processes (pp. 271–317). New York: Academic Press.

Hanel, P. H., & Vione, K. C. (2016). Do Student Samples Provide an Accurate Estimate of the General Public? *PloS one, 11*(12), e0168354. https://doi.org/10.1371/journal.pone.0168354

Hertwig, R, Pleskac, T., & Pachur, T. (2019). Reckoning with Uncertainty: Our Program of Research. In R. Hertwig, T. Pleskac, & T. Pachur (Eds.) *Taming Uncertainty*. Boston, MA: MIT Press.

Hilton, D. J. (2011). Linguistic polarity, outcome framing, and the structure of decision making: A pragmatic approach. In G. Keren (Ed.), *Perspectives on Framing*. (pp. 135-155). New York, NY, US: Psychology Press.

Hogarth, R. M., Lejarraga, T., & Soyer, E. (2015). The two settings of kind and wicked learning environments. *Current Directions in Psychological Science, 24*(5), 379-385. https://doi.org/10.1177/0963721415591878

Jekich, B. M., Drake, B. D., Nacht, J. Y., Nichols, A., Ginde, A. A. , & Davis, C. B. (2016). Avalanche fatalities in the United States: A change in demographics. *Wilderness & Environmental Medicine, 27*(1), 46–52. https://doi.org/10.1016/j.wem.2015.11.004

Johnson, J., Mannberg, A., Hendrikx, J., Hetland, A., & Stephensen, M. (2020). Rethinking the Heuristic Traps Paradigm in Avalanche Education: Past, Present and Future. *Cogent Social Sciences, 6*(1). https://doi.org/10.1080/23311886.2020.1807111

Juslin, P. (1994). The Overconfidence Phenomenon as a Consequence of Informal Experimenter-Guided Selection of Almanac Items. *Organizational Behavior and Human Decision Processes, 57*(2), 226-246. https://doi.org/10.1006/obhd.1994.1013

Juslin, P., & Montgomery, H. (Eds.). (2007) *Judgment and decision making: Neo-Brunswikian and process-tracing approaches*. New York, NY: Psychology Press

Juslin, P., Winman, A., & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard–easy effect. *Psychological Review, 107*(2), 384-396. https://doi.org/10.1037/0033-295X.107.2.384

Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition, 11*(2), 143-157. https://doi.org/10.1016/0010-0277(82)90023-3

King, J., & Slovic, P. (2014). The affect heuristic in early judgments of product innovations. *Journal of Consumer Behaviour, 13*(6), 411-428. https://doi.org/10.1002/cb.1491

Kirchler, E., Maciejovsky, B., & Weber, M. (2005). Framing Effects, Selective Information, and Market Behavior: An Experimental Analysis. *Journal of Behavioral Finance, 6*(2), 90-100. https://doi.org/10.1207/s15427579jpfm0602_4

Koehler, D. J., Brenner, L., & Griffin, D. (2002). The Calibration of Expert Judgment: Heuristics and Biases Beyond the Laboratory. In D. Griffin, D. Kahneman, & T. Gilovich (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 686-715). Cambridge: Cambridge University Press.

Koriat, A. (2018). When reality is out of focus: Can people tell whether their beliefs and judgments are correct or wrong? *Journal of Experimental Psychology: General, 147*(5), 613-631. https://doi.org/10.1037/xge0000397

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*(3), 490-517. https://doi.org/10.1037/0033-295X.103.3.490

Kozyreva, A., & Hertwig, R. (2021). The interpretation of uncertainty in ecological rationality. *Synthese, 198*(2), 1517-1547. https://doi.org/10.1007/s11229-019-02140-w

Kühberger, A. (1998). The Influence of Framing on Risky Decisions: A Meta-analysis. *Organizational Behavior and Human Decision Processes, 75*(1), 23-55. https://doi.org/10.1006/obhd.1998.2781

Kühberger, A. (2017). Framing. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory, 2nd ed.* (pp. 79-98). New York, NY, US: Routledge.

Landrø, M., Hetland, A., Engeset, R. V., & Pfuhl, G. (2020a). Avalanche decision-making frameworks: Factors and methods used by experts. *Cold Regions Science and Technology, 170*, 102897. https://doi.org/10.1016/j.coldregions.2019.102897

Landrø, M., Pfuhl, G., Engeset, R., Jackson, M., & Hetland, A. (2020b). Avalanche decision-making frameworks: Classification and description of underlying factors. *Cold Regions Science and Technology, 169*, 102903. https://doi.org/10.1016/j.coldregions.2019.102903

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and Decision Making. Annual Review of Psychology, 66(1), 799-823. https://doi.org/10.1146/annurev-psych-010213-115043

Levin, I. P., & Gaeth, G. J. (1988). How Consumers are Affected by the Framing of Attribute Information Before and After Consuming the Product. *Journal of Consumer Research, 15*(3), 374-378. https://doi.org/10.1086/209174

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All Frames Are Not Created Equal: A Typology and Critical Analysis of Framing Effects. *Organizational Behavior and Human Decision Processes, 76*(2), 149-188. https://doi.org/10.1006/obhd.1998.2804

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin, 127*(2), 267-286. https://doi.org/10.1037/0033-2909.127.2.267

Mannberg, A., Hendrikx, J., & Johnson, J. (2020). Risky positioning – social aspirations and risk-taking behaviour in avalanche terrain. *Leisure Studies*, 1-18. https://doi.org/10.1080/02614367.2020.1831046

Mannberg, A., Hendrikx, J., Landrø, M., & Ahrland Stefan, M. (2018). Who's at risk in the backcountry? Effects of individual characteristics on hypothetical terrain choices. *Journal of environmental psychology*, *59*, 46-53. https://doi.org/10.1016/j.jenvp.2018.08.004

Maule, J., & Villejoubert, G. (2007). What lies beneath: Reframing framing effects. *Thinking & Reasoning, 13*(1), 25-44. https://doi.org/10.1080/13546780600872585

McCammon, I., & Hägeli, P. (2007). An evaluation of rule-based decision tools for travel in avalanche terrain. *Cold Regions Science and Technology, 47*(1), 193-206. https://doi.org/10.1016/j.coldregions.2006.08.007

McClung, D. & Schaerer, P.A. (2006). *The Avalanche Handbook*. Seattle, WA: The Mountaineers Books.

Meyniel, F., Schlunegger, D., & Dehaene, S. (2015). The Sense of Confidence during Probabilistic Learning: A Normative Account. *PLOS Computational Biology, 11*(6), e1004305. https://doi.org/10.1371/journal.pcbi.1004305

Morrow, D. G. (2018). Publishing papers that matter. *Journal of Experimental Psychology: Applied, 24*(1), 1-2. http://dx.doi.org/10.1037/xap0000141

Newell, B. R. (2013). Judgment under uncertainty. In D. Reisberg (Ed.) *The Oxford Handbook of Cognitive Psychology.* (pp. 603-617). New York, NY, US: Oxford University Press.

Niedermeier, M.; Gatterer, H.; Pocecco, E.; Frühauf, A.; Faulhaber, M.; Menz, V.; Burtscher, J.; Posch, M.; Ruedl, G.; & Burtscher, M. (2020). Mortality in Different Mountain Sports

Activities Primarily Practiced in the Winter Season—A Narrative Review. *International Journal of Environmental Research and Public Health, 17(259)*. https://doi.org/10.3390/ijerph17010259

Oskarsson, A. T., Van Boven, L., McClelland, G. H., & Hastie, R. (2009). What's next? Judging sequences of binary events. *Psychological Bulletin, 135*(2), 262–285. https://doi.org/10.1037/a0014821

Payne, J.W., Sagara, N., Shu, S.B. *et al.* (2013) Life expectancy as a constructed belief: Evidence of a live-to or die-by framing effect. *Journal of Risk and Uncertainty*, 46**,** 27–50. https://doi.org/10.1007/s11166-012-9158-0

Peters, E., Hart, P. S., & Fraenkel, L. (2011). Informing patients: the influence of numeracy, framing, and format of side effect information on risk perceptions. *Medical Decision Making, 31*(3), 432-436. https://doi.org/10.1177/0272989x10391672

Petrusic, W. M., & Baranski, J. V. (2003). Judging confidence influences decision processing in comparative judgments. *Psychonomic Bulletin & Review, 10*(1), 177-183. https://doi.org/10.3758/BF03196482

Pham, M. T. (1998). Representativeness, Relevance, and the Use of Feelings in Decision Making. *Journal of Consumer Research, 25*(2), 144-159. https://doi.org/10.1086/209532

Piñon, A., & Gambara, H. (2005). A meta-analytic review of framing effect: Risky, attribute and goal framing. *Psicothema, 17*(2), 325-331.

Rieskamp, J., & Otto, P. E. (2006). SSL: A Theory of How People Learn to Select Strategies. *Journal of Experimental Psychology: General, 135*(2), 207-236. https://doi.org/10.1037/0096-3445.135.2.207

Rothman, A. J., & Salovey, P. (1997). Shaping perceptions to motivate healthy behavior: The role of message framing. *Psychological Bulletin, 121*(1), 3-19. https://doi.org/10.1037/0033-2909.121.1.3

Sanchez, C., & Dunning, D. (2018). Overconfidence among beginners: Is a little learning a dangerous thing? *Journal of Personality and Social Psychology, 114*(1), 10-28. https://doi.org/10.1037/pspa0000102

Sanchez, C., & Dunning, D. (2020). Decision fluency and overconfidence among beginners. *Decision, 7*(3), 225-237. https://doi.org/10.1037/dec0000122

Scheufele, D.A. & Tewksbury, D. (2007), Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models. *Journal of Communication, 57*(1), 9-20. https://doi.org/10.1111/j.0021-9916.2007.00326.x

Schurz, G., & Hertwig, R. (2019). Cognitive success: A consequentialist account of rationality in cognition. *Topics in Cognitive Science, 11*(1), 7–36. https://doi.org/10.1111/tops.12410

Schwarz, N. (2007). Attitude Construction: Evaluation in Context. *Social Cognition, 25*(5), 638-656. https://doi.org/10.1521/soco.2007.25.5.638

Schwarz, N. (2012). Feelings-as-Information Theory. In P. A. M. Van Lange, A. W. Kruglanski, & T. Higgins (Eds.), *Handbook of Theories of Social Psychology: Volume 1* (pp. 289-308). London: Sage Publications Ltd.

Schweizer, J., Lütschg, M. (2000). Measurements of human-triggered Avalanches from the Swiss Alps. Proceedings from International Snow Science Workshop, Big Sky, Montana, U.S.A.

Simenhois, R., & Birkeland, K. W. (2009). The Extended Column Test: Test effectiveness, spatial variability, and comparison with the Propagation Saw Test. *Cold Regions Science and Technology, 59*(2), 210-216. https://doi.org/10.1016/j.coldregions.2009.04.001

Slovic, P. (1987). Perception of risk. *Science, 236*(4799), 280. https://doi.org/10.1126/science.3563507

Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). The Affect Heuristic. In D. Griffin, D. Kahneman, & T. Gilovich (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 397-420). Cambridge: Cambridge University Press.

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2004). Risk as Analysis and Risk as Feelings: Some Thoughts about Affect, Reason, Risk, and Rationality. *Risk Analysis, 24*(2), 311-322. https://doi.org/10.1111/j.0272-4332.2004.00433.x

Slovic, P., Peters, E., Finucane, M. L., & MacGregor, D. G. (2005). Affect, risk, and decision making. *Health Psychology, 24*(4, Suppl), S35-S40. https://doi.org/10.1037/0278-6133.24.4.S35

Speekenbrink, M., & Shanks, D. R. (2013). Decision making. In D. Reisberg (Ed.) *The Oxford Handbook of Cognitive Psychology.* (pp. 682-703). New York, NY, US: Oxford University Press.

Statham, G., Haegeli, P., Greene, E., Birkeland, K., Israelson, C., Tremper, B., Stethem, C., McMahon, B., White, B., & Kelly, J. (2018). A conceptual model of avalanche hazard. *Natural Hazards, 90*(2), 663–691. https://doi.org/10.1007/s11069-017-3070-5

Stephensen, M. B., Landrø, M., & Hendrikx, J. (2021) Should we judge danger or safety in avalanche terrain? *The Avalanche Journal*, 126, 18-19.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* Yale University Press.

Tremper, Bruce. (2008). *Staying Safe in Avalanche Terrain*. London: Bâton Wicks.

Ülkümen, G., Fox, C. R., & Malle, B. F. (2016). Two dimensions of subjective uncertainty: Clues from natural language. *Journal of Experimental Psychology: General, 145*(10), 1280-1297. https://doi.org10.1037/xge0000202

Weber, M., Keppe, H.-J., & Meyer-Delius, G. (2000). The impact of endowment framing on market prices — an experimental analysis. *Journal of Economic Behavior & Organization, 41*(2), 159-176. https://doi.org/10.1016/S0167-2681(99)00092-X

Wilson, R. S., & Arvai, J. L. (2006). When Less is More: How Affect Influences Preferences When Comparing Low and High-risk Options. *Journal of Risk Research, 9*(2), 165-178. https://doi.org/10.1080/13669870500419503

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1594), 1310-1321. https://doi.org10.1098/rstb.2011.0416

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35*(2), 151-175. https://doi.org/10.1037/0003-066X.35.2.151

Zweifel, B., & Haegeli, P. (2014). A qualitative analysis of group formation, leadership and decision making in recreation groups traveling in avalanche terrain. *Journal of Outdoor Recreation and Tourism, 5-6*, 17-26. https://doi.org/10.1016/j.jort.2014.03.001

Paper 1

PREPUBLICATION VERSION

Should I Judge Safety or Danger? Perceived Risk Depends on the Question Frame

Matthew B. Stephensen[1], Christin Schulze[2,1], Markus Landrø[1,3], Jordy Hendrikx[4], and Audun Hetland[1]

[1]UiT The Arctic University of Norway, Tromsø, Norway
[2]Max Planck Institute for Human Development, Berlin, Germany
[3]Norwegian Water Resources and Energy Directorate, Oslo, Norway
[4]Montana State University, Bozeman, USA

Open science statement: Data, analysis scripts, and study materials are available at https://osf.io/sknxf/, https://doi.org/10.17605/OSF.IO/SKNXF

The copy of record of this article is available at:

Stephensen, M. B., Schulze, C., Landrø, M., Hendrikx, J., & Hetland, A. (2021). Should I judge safety or danger? Perceived risk depends on the question frame. *Journal of Experimental Psychology: Applied*. https://doi.org/10.1037/xap0000354

Abstract

Linguistic polarity is a natural characteristic of judgments: Is that situation safe/dangerous? How difficult/easy was the task? Is that politician honest/dishonest? Across six studies (N = 1599), we tested how the qualitative frame of the question eliciting a risk judgment influenced risk perception and behavior intention. Using a series of hypothetical scenarios of skiing in avalanche terrain, experienced backcountry skiers judged either *how safe* or *how dangerous* each scenario was and indicated whether they would ski the scenario. Phrasing risk judgments in terms of safety elicited lower judged safety values, which in turn resulted in a lower likelihood of intending to ski the slope. The frame "safe" did not evoke a more positive assessment than the frame "danger" as might be expected under a valence-consistent or communication-driven framing effect. This seemingly paradoxical direction of the effect suggests that the question frame directed attention in a way that guided selective information sampling. Uncertainty was not required for this effect as it was observed when judging objectively safe, uncertain, and dangerous scenarios. These findings advance our theoretical understanding of framing effects and can inform the development of practices that harness question framing for applied risk perception and communication.

Keywords: Framing effect; risk perception; judgment and decision making; attention; avalanche terrain

Public Significance Statement

This study demonstrates that risk judgments framed in terms of safety (*How safe is it?*) result in more cautious, conservative judgments than when framed in terms of danger (*How dangerous is it?*), irrespective of the objective risk of the judged object. These findings advance our understanding of the framing effect while illustrating its particular relevance for applied risk perception practices and for public hazard forecasting and information communication strategies.

Should I Judge Safety or Danger? Perceived Risk Depends on the Question Frame

Framing is a ubiquitous psychological phenomenon that holds much promise for applied risk communication and risk perception. A wealth of empirical findings have shown that the way in which a problem, situation, object, event, or goal is described affects how people respond to or evaluate it (for reviews see e.g., Kühberger, 1998, 2017; Levin et al., 1998; Maule & Villejoubert, 2007; Piñon & Gambara, 2005). Framing effects are traditionally regarded as a deviation from rational behavior (Kahneman & Tversky, 1984; Tversky & Kahneman, 1981). Recognizing the potential problems that framing effects can cause, researchers have investigated numerous methods for debiasing or diminishing framing effects (Almashat et al., 2008; Garcia-Retamero & Dhami, 2013; Sieck & Yates, 1997; Simon et al, 2004). Yet not all instances of framing effects are considered problematic, harmful, or disadvantageous. The purposeful use of framing to promote specific judgments and decisions is an established practice in several fields such as media and communications (Block & Keller, 1995; Chong & Druckman, 2007; Scheufele & Tewksbury, 2007) and marketing (Biswas, 2009; Biswas & Grau, 2008; Levin & Gaeth, 1988). Might the framing effect be similarly harnessed to boost people's natural decision making competencies in order to ensure better, safer risk perceptions? If so, the framing effect has the potential for a real and tangible impact on how people judge risk and, ultimately, how they behave under risky conditions.

In this article, we employ decision making in avalanche terrain as an exemplary case for investigating risk judgments and behavior in domains of applied risk perception and risk communication. Decision making in avalanche terrain represents a paradigmatic case of applied risk judgments and decisions. Avalanche terrain is a highly complex and uncertain decision environment in which judgment errors can result in severe injury or death. Moreover, it is an environment where poor decisions are frequently marked by invalid feedback (cf. Hogarth et al., 2015). A mismatch between perceived risk and reality can therefore be present for even the most experienced decision makers. Nonetheless, skiing in avalanche terrain is an increasingly popular but largely unregulated activity in which people voluntarily engage, most often for purely recreational purposes. Decision making in avalanche terrain provides a paradigmatic case for investigating decisions under real-life uncertainty among an experienced, highly motivated population. Yet the relevance of this research is not limited to backcountry skiing. Frame selection and the strategic phrasing of risk judgments are relevant for a range of disciplines and contexts including police, fire and other emergency services, military operations, the finance sector, work in volatile natural environments such as fisheries, illicit activities such as recreational drug use, and for medical diagnoses and treatment decisions.

Decision making in avalanche terrain offers a convenient exemplary case for investigating the effect of framing on risk perception under varying degrees of uncertainty.

One type of framing that appears to be highly relevant for applied risk perception and risk communication is the question frame. Judgments are commonly framed in terms of a single dimension of an integral qualitative attribute of what is judged. For example, "How *bad* is the situation?", "How *expensive* is a product?", or "How *difficult is* an activity?" (as opposed to how good, inexpensive, or easy). Polarity is a natural characteristic of language that constrains and thereby defines the formulation of many judgments to a single dimension of a bipolar attribute (Hilton, 2011). Decision makers therefore naturally employ a single dimension of a qualitative reference such as *safe* or *dangerous* to frame a risk judgment.

Using hypothetical scenarios of backcountry skiing in avalanche terrain across six studies (N = 1599), we examined how recreational backcountry skiers evaluated these scenarios when asked to judge *how safe* versus *how dangerous* they are. We also examined how the question frames influenced participants' decision to ski the scenarios. We tested the effect of the question frames on risk perception and behavior intention when judging scenarios of different objective risk levels: scenarios of uncertain risk (Studies 1-6), safe scenarios (Studies 5-6), and dangerous scenarios (Studies 5-6). Based on the existing framing literature, we can expect to find an effect when framing a risk judgment in terms of safety or danger. Beyond that, however, prior research has not established the direction of that question framing effect, the prevalence of that effect under different conditions of objective risk, or the association of that framing effect with behavior intention. How does a qualitative reference frame influence perceived risk? Which of the two frames – safe or dangerous – evokes a more conservative, cautious risk judgment? Is uncertainty a requirement for an effect or does it persist in the face of credible evidence of safety or danger? How does the question framing effect influence behavior, which is ultimately what exposes a person to risk? Different accounts of the cognitive processes underlying framing effects make conflicting predictions about the direction and prevalence of an effect evoked by a question frame. These issues must be resolved in order to determine if it might be possible to harness question framing to promote safer risk judgments and decisions.

Divergent Accounts of Framing and the Direction of the Predicted Effect

Framing effects are diverse and inspire broad theoretical and applied interest in psychology, economics, political science, health sciences, and beyond. Consequently, a variety of heterogeneous phenomena that evoke differing cognitive processes and divergent effects are classified as framing (for reviews, see e.g., Chong & Druckman, 2007; Kühberger, 1998, 2017;

Levin et al., 1998; Maule & Villejoubert, 2007). Research on framing in psychology and economics has primarily focused on situations in which different but logically equivalent descriptions lead to different preferences or evaluations, highlighting a violation of the economic principle of invariance (e.g., Tversky & Kahneman, 1981). For example, evaluations tend to be more favorable when a product is described as 90% fat-free than when described as containing 10% fat (e.g., Levin, 1987). This research has traditionally used risky choice framing in which the probabilities of the choice options are differently described (e.g., the classic "Asian disease problem"; Tversky & Kahneman, 1981), and attribute framing in which a single attribute of the target of judgment is differently described (Levin et al., 1998; Piñon & Gambara, 2005). Manipulating the qualitative attribute that frames the question eliciting a judgment (i.e., question framing) has received much less attention in this line of research and has produced mixed results (see e.g., Payne et al., 2013; Comerford & Robinson, 2017).

Research in the fields of political science and health sciences have applied a broader conception of framing in which emphasis on different aspects of an issue leads to different opinions, without necessitating logical equivalence between frames (e.g., Cacciatore et al., 2016; Chong & Druckman, 2007; Druckman, 2001; Bui et al., 2015; Nelson et al., 1997). For example, when asked about government funded financial assistance for people in need, political opinions tend to be more favorable toward government spending if preempted with an emphasis on humanitarian aspects rather than government expenditures (Druckmann, 2001). This broader conception of framing effects is also relevant to survey studies (see Bruine de Bruin, 2011), in which questions with presumed synonyms can elicit inconsistent responses (e.g., Bruine de Bruin et al., 2012) and questions with presumed antonyms can fail to communicate polar opposites (e.g., Holleman, 2006).

The cognitive processes that a frame activates and the direction of the resulting framing effect likely depend on the form and domain in which the frame is achieved. Consequently, several cognitive, communicative, and attentional processes have been proposed for framing effects achieved inside and outside the lab (for a review, see Keren, 2011). These different accounts of framing make somewhat conflicting predictions about how the question frame we tested might affect people's judgment. Does a question about "safety" evoke a more reassuring assessment than a question about "danger" because it triggers a positive connotation or conveys the assumption of safety in the question? Alternatively, does a question about "safety" direct a decision maker's attention to the sparsity of evidence of safety under conditions of uncertainty, thereby prompting a more conservative evaluation? These different theoretical accounts of framing were established and predominantly tested based on risky choice framing and attribute

framing research. The extent to which they describe and can predict a question framing effect is presently unclear (see Comerford & Robinson, 2017). By testing the direction of the framing effect evoked when risk judgments are framed in terms of safety and danger, we can infer the cognitive processes that are activated. Because the success of any effort to strategically harness framing in applied contexts depends on correctly matching the type of frame – and the cognitive process that it activates – with the objective for its application, we next review these different accounts of framing in the context of our study.

Valence-driven account of framing effects

One prominent cognitive account posits that framing information in either a positive or a negative way evokes a *valence-consistent* association that influences the selection and encoding of information about the target(s) of judgment (for reviews, Keren, 2011; Levin et al., 1998; Piñon & Gambara, 2005; Teigen, 2015). For example, positively framing an action (75% chance of success) evokes positive associations resulting in more favorable judgments of that action than does negatively framing the same action (25% chance of failure), despite the two frames being logically equivalent. Investigating the valence account of framing in the context of question frames, Payne and colleagues (2013) found that life expectancy predictions were longer when judging the probability to "*live to*" a certain age than when judging the probability to "*die by*" that age. Subjective probability judgments about longevity and verbal protocols both indicated that the "*live to*" frame evoked more positive thoughts than the "*die by*" frame did. For our context of risk judgments, assuming that the quality *safe* has a positive valence and the quality *dangerous* has a negative valence (Hedger et al., 2016), then according to the valence account of framing, a risk judgment phrased as *How safe is it?* should evoke positive associations that may result in higher judged safety than a risk judgment phrased as *How dangerous is it?* However, other empirical findings cast doubt on whether valence consistent associations adequately account for question framing effects. Although Comerford and Robinson (2017) replicated the results of Payne et al (2013), they also found that the response format influenced the direction of the question framing effect and, we can assume, the underlying cognitive processes. When decision makers reported life expectancy as a point estimate in response to a framed statement "*I expect to live to/die by age…*", life expectancy was longer under the "*die by*" frame. These surprisingly contradictory results highlight the continuing uncertainty about the cognitive processes activated by attribute framing of questions and the direction of the evoked effect.

Communicative accounts of framing effects

Another influential account of framing holds that the pragmatics of language and communication contribute to framing effects (Hilton, 2011; McKenzie, 2004; McKenzie & Nelson, 2003; Sher & McKenzie, 2006). A frame is typically selected by a source (i.e., a speaker) communicating information about the target(s) of judgment. Importantly, the speaker's choice of frame and the listener's inferences about that choice are not arbitrary. Consequently, the choice of frame "leaks" implicit information about a target beyond what is explicitly stated. In this way, logically equivalent frames might not be informationally equivalent because the choice of frame conveys judgment-relevant information, notably the communicator's perspective on the target of judgment. Differently framed questions in survey research are typically also understood according to such a communicative theoretical framework that regards the interaction between researcher and survey respondent as a form of communication subject to the rules of everyday conversation (Grice, 1975). In our study, the differently framed question that elicits the risk judgment could be interpreted as leaking different information about the communicator's (i.e., the researcher's) perception of the target of judgment. When asked the question "*How safe is it?*" the decision maker may plausibly assume that the communicator asking the question perceives the target of judgment to be safe. Otherwise, the communicator would have asked "*How dangerous is it?*" if the target was perceived to be dangerous. Although the cognitive process believed to underlie this framing effect differs from the valence account, the communication account of framing might similarly predict that the term *safe* in the question would elicit judgments of higher safety than would the term *dangerous* in the question.

Attentional accounts of framing

A final account of framing highlights attentional processes. These assert that a frame cues the cognitive system to direct attentional resources toward a certain perspective on the target(s) of judgment while suppressing attention toward alternative perspectives (e.g., Keren, 2011; Teigen, 2015). Judging a target involves cognitive processes that operate in relative terms. All judgments are relative to a reference that is the focus of attention, and the nature and location of that reference influence the judgment (Keren, 2011). Research in psychology using numerically framed single-bound probability judgments found that a frame defines a descriptive state (e.g., more than 85%) as the provisional reference point for the judgment (Hohle & Teigen, 2018; Teigen et al., 2007). The phrasing of the judgment task directs the decision maker's attention toward evaluating whether the target of judgment meets or fulfills that descriptive state, and the decision maker samples different information according to the

perspective or reference defined by the frame. For example, a weather forecast predicting that the chance of rain is "greater than 60%" guides attention toward the occurrence whereas a forecast of "less than 70%" guides attention toward the non-occurrence of rain. In that way, a decision maker's judgment in response to a frame is analogous with selectively testing a focal hypothesis defined by that frame (Teigen, 2015). Research in political science on emphasis framing similarly asserts that framing focuses attention on a certain perspective or issue, and consequently that focal issue will have greater weight during the judgment process because of its increased accessibility and applicability (Block & Keller, 1995; Chong & Druckman, 2007; Scheufele & Tewksbury, 2007).

From the perspective of the attention account of framing, we might expect the use of the term *safe* in the question to focus the decision maker's attention on sampling evidence to test the hypothesis that the scenario is indeed safe. Yet when uncertainty is high and definitive evidence is lacking, the decision maker might be expected to conclude that the scenario is not safe because of the insufficiency of evidence in support of that hypothesis. The term dangerous in the question would similarly focus the decision maker's attention on searching for evidence of danger. When no definitive evidence of danger is found, the decision maker would conclude that the scenario is not dangerous because of the insufficiency of evidence in support of that hypothesis. According to the attention account of framing, one would expect judging how safe a situation is to elicit more cautious, more conservative judgments (i.e., *lower* judged safety ratings) than judging how dangerous a situation is under conditions of uncertainty. Conversely, one would expect judging how dangerous a situation is to elicit less cautious, less conservative judgments (i.e., *lower* danger rating or, conversely, *higher* judged safety ratings) than judging how safe a situation is. Importantly, however, it is unclear whether uncertainty is a necessary condition for the cognitive process described by the attention account to produce a framing effect. One could reasonably assume that there would be no question framing effect when encountering sufficient evidence to establish objective safety under the safe frame, or when encountering sufficient evidence to establish objective danger under the danger frame. Although the question frame likely directs evidence sampling under all conditions of (un)certainty, this might only result in a framing effect under conditions of objective uncertainty.

Summary and Research Hypotheses

Three influential accounts of framing make opposing predictions about how the question frames we tested might affect the perceived risk of scenarios of backcountry skiing in avalanche terrain. It is important to note, however, that none of these accounts were developed

in the context of qualitatively framing the question that elicits a judgment. Nevertheless, all of these accounts have been used to explain different types of question frames, as reviewed above. Our study differs from previous research investigating these accounts of framing in that we presented participants with highly uncertain visual scenarios in addition to verbal question frames. Although both the valence account and the communication account of framing might predict that judging how safe a situation is would initially elicit higher judged safety, neither account makes strong predictions about how participants subsequently search for information. It is also possible that several or all of the cognitive processes proposed by the different accounts of framing are activated simultaneously or sequentially upon encountering the question frame, in which case we might expect a mixture of effects or no overall effect of framing whatsoever. We test the effect of question framing under conditions of objective uncertainty, safety, and danger in order to ascertain the direction(s) of the effect(s), and thereby infer the cognitive processes activated by a risk judgment framed in terms of safety or danger.

Based on our review of the abovementioned accounts of framing, we predicted in Studies 1 to 4 that the framing of the question that elicited the risk judgment would influence the judged safety of the scenarios, all of which were of uncertain risk. In line with previous work on valence-consistent question framing (e.g., Payne et al., 2013), in Studies 1 and 2 we predicted that a risk judgment phrased as "*How safe is it?*" would elicit higher judged safety than a risk judgment phrased as "*How dangerous is it?*" When the results of Studies 1 and 2 did not confirm this valence-consistent hypothesis, we updated the direction of the predicted effect in Studies 3 and 4 according to the attention account of framing (see, e.g., Keren, 2011). We predicted that the question "*How safe is it?*" would elicit lower judged safety than the question "*How dangerous is it?*" In Studies 5 and 6, we investigated whether the direction of the question framing effect is consistent when judging scenarios of different objective risk levels. We tested two competing hypotheses: 1) the experimental condition would have the same effect on the judged safety of all scenario categories, versus 2) the experimental condition would only influence the judged safety of scenarios of uncertain risk and there would be negligible effects for objectively safe and dangerous scenarios. Finally, we also predicted for all six studies that the variable judged safety would predict behavior intention. In what follows, we report the methods of data collection and analysis that were common for all six studies, and specify any aspects that were unique to any study. We then report the results of each of the six studies.

Methods of Data Collection and Analysis for Studies 1 to 6

Across six studies (N = 1599), we tested how risk perceptions and decisions are influenced by the qualitative frame of the question that elicits the risk judgment (i.e., the question frame). Using a series of hypothetical scenarios involving backcountry skiing in avalanche terrain, participants judged either *how safe* or *how dangerous* each scenario was and indicated whether they would ski the scenario. We confirm that at the time of writing, the six studies reported in this article are all the studies we conducted on the effect of question framing on risk judgments and decisions. We report all measurements assessed and all manipulations implemented in each study. The studies were approved by the Department of Psychology Research Ethics Committee, UiT The Arctic University of Norway. All studies except Study 4 were pre-registered. The pre-registrations, data, R script for data processing and analysis, and the scenarios used in the studies are available on the Open Science Framework (https://osf.io/sknxf/).

Participants. Table 1 presents the sample size and information about the sample for each of the six studies. Due to a technical failure, we were unable to collect data on participant age, gender, or experience measures in Study 1. Each of the six studies was conducted during a public seminar on safety and decision-making for backcountry skiers in avalanche terrain, with the seminar audience members as participants. Each seminar was at a different location in Norway during the winter of 2019 to 2020. The audience members at the six seminars were recreational backcountry skiers with varying degrees of experience judging avalanche risk during ski tours in avalanche terrain. Consequently, there was a self-selection for experienced participants in all six studies. Table 2 presents descriptive statistics on participants' experience with the judgment tasks and the activity presented in the scenarios. The seminar in which we conducted Study 1 had a nominal entry fee of NOK 50 (approximately 6 USD) whereas the other five seminars were free to attend. Participation was voluntary and all participants indicated their informed consent to participate. All six studies were conducted in Norwegian.

We did not conduct a priori power analysis to determine target sample size for any study. The audience size at a seminar determined the possible number of participants in the study conducted at that event. We recruited as many participants as possible during each seminar and did not continue data collection for the respective study beyond that seminar, but otherwise had no control over the final sample size of each study. We set a minimum sample size for each study of approximately 60% of the anticipated audience size for that seminar. Participation exceeded 60% of the actual audience size for all six studies and we met the estimated minimum number of participants for Studies 1 to 5. Although more than 60% of the

audience participated in Study 6, we overestimated the expected audience size and did not meet the minimum number of expected participants.

Table 1

Participants in Studies 1 to 6, including sample sizes, division of participants by experimental condition, and data on participants' gender and age. Gender self-identification categories are male (M), female (F), other (O), withheld (W), and data not available (NA). Gender and age data are unavailable for Study 1.

Study	N	Experimental condition		Gender self-identification					M age	SD age	Range age
		Safe	Danger	M	F	O	W	NA			
1	735	351	384	-	-	-	-	735	-	-	-
2	197	102	95	131	53	2	1	10	37	10.95	18-62
3	255	132	122	156	99	0	0	0	38	11.40	18-65
4	173	86	87	90	81	2	0	0	37	11.79	18-73
5	168	66	102	94	74	0	0	0	32	10.17	18-67
6	71	32	39	35	34	2	0	0	29	8.91	19-53

Table 2

Participant experience with the judgment tasks for Studies 1 to 6, including average number of years skiing in avalanche terrain (M_{Years}, response scale from 0 to 40 years), average number of days in avalanche terrain per season (M_{Days}, response scale from 0 to 100 days), and median level of self-reported avalanche safety training ($M_{Training}$, reported on 7-point scale with "1 – None" and "7 – Expert level qualifications" at the scale ends).

Study	N	M_{Years}	M_{Days}	$M_{Training}$
1	735	-	-	-
2	197	8.51	12.02	3
3	255	11.77	14.38	2
4	173	9.54	15.03	2
5	168	7.87	8.36	3
6	71	8.10	28.17	3

Design. All six studies used the same experimental design, measures, and procedures described here. We used a between-subject design with two experimental conditions—the Safe Group or the Danger Group—for the qualitative attribute that framed the risk judgments. All judgment tasks were programmed in Qualtrics. One seminar leader conducted Studies 1, 2 and 6, another seminar leader conducted Studies 3 and 5, and a third seminar leader conducted Study 4. Each seminar began with a presentation of the information resources that are publicly available online from the Norwegian Avalanche Warning Service. After that, the seminar leader announced the study explaining that researchers were investigating information use for avalanche risk judgments. The seminar leader then projected the link to the online study for participants to access via their internet connected mobile devices. Upon accessing the study, participants were randomly assigned by the software to one of the two experimental conditions after indicating their informed consent to participate.

Materials. Participants judged a series of hypothetical scenarios of backcountry skiing in avalanche terrain. Each scenario consisted of a photograph of a person skiing a snow-covered slope (one scenario photo included three people). We hold the rights of use for all photos. In the upper right corner of each photograph were icons indicating the slope angle, the prevailing regional avalanche problem(s), and the forecasted regional avalanche danger level (5-point scale) for that scenario. The icons used are standardized icons defined by the European Avalanche Warning Services (EAWS) and used by the Norwegian Avalanche Warning Service in daily regional avalanche danger forecasts throughout the country. These icons provide valuable objective information for judging the degree of risk.

We developed three categories of scenarios: scenarios of uncertain risk, safe scenarios, and dangerous scenarios, as defined by avalanche experts. An avalanche expert selected photographs from a personal library and assigned icons to those photographs to create scenarios of differing risk level. The combination of visual evidence in the photograph and the information conveyed by the icons established the objective risk level of each scenario. Uncertain scenarios lacked definitive evidence of safety or danger because the available evidence in the photo and the icons were ambiguous and/or conflicting. For example, although the icons on the steepness of the slope and the prevailing avalanche problems together indicate an increased probability of an avalanche, the icon indicating a danger level of two (on a five-point scale) and the terrain features depicted in the photograph indicate a reduced probability of avalanche. The risk level of that scenario would be objectively uncertain given the conflicting evidence. By contrast, the photographs and icons in the safe and dangerous scenarios conveyed sufficient evidence to ascertain the objective safety or danger of the scenario. For example,

although an icon indicates the prevailing regional avalanche problem, no signs indicative of the presence of that problem are evident in the photograph. Additional icons in that scenario indicate a low regional danger forecast and a low slope angle. That scenario is objectively safe because all the evidence align to indicate that an avalanche is extremely unlikely. The objective uncertainty of the uncertain scenarios, the objective safety of the safe scenarios, and objective danger of the dangerous scenarios were established by the avalanche expert who designed the scenarios, and independently confirmed by a second avalanche expert. All scenarios used in the six studies are available on the Open Science Framework at https://osf.io/sknxf/.

Participants in Studies 1 to 4 judged six uncertain scenarios. We used the same six uncertain scenarios for all four studies, changing their order of presentation between studies to account for any possible order effect. Participants in Studies 5 and 6 judged ten scenarios: four scenarios of uncertain risk, three safe scenarios, and three dangerous scenarios. We used the same set of ten scenarios for both studies. The four uncertain scenarios used in Studies 5 and 6 were selected from among the six uncertain scenarios previously used in Studies 1 to 4. The three safe scenarios and the three dangerous scenarios used in Studies 5 and 6 were new. We anticipated the possibility of an effect from the order in which the categories of scenarios were judged. We therefore reversed the order of presentation between Studies 5 and 6. In Study 5, we first presented the three dangerous scenarios followed by the four uncertain scenarios, and lastly the three safe scenarios. In Study 6, we first presented the three safe scenarios, then the four uncertain scenarios, and finally the three dangerous scenarios. Although we reversed the order of the scenario categories in Study 6, the order of the scenarios within each category was the same in both studies.

Measures. The scenarios were sequentially projected onto the auditorium screen for all participants in the audience to see. All questions were displayed exclusively in Qualtrics on participants' personal mobile devices. Participants in the Safe Group judged the scenarios by answering the question "*How safe is it?*" responding on a 7-point scale labeled "*Not at all safe*" and "*Completely safe*" at the extreme points. Participants in the Danger Group judged the same scenarios by answering the question "*How dangerous is it?*" responding on a 7-point scale labeled "*Not at all dangerous*" and "*Completely dangerous*" at the extreme points. Upon completing each risk judgment, participants in both experimental groups were asked the question "*Would you ski this slope?*" with the three response options "*No*", "*I cannot say*" or "*Yes*". Participants had approximately one minute per scenario to answer the risk judgment and the behavior intention questions before the next scenario was projected. Participants were instructed not to discuss with their neighbors during the study and the seminar leaders

confirmed that all auditoriums were silent during data collection. Participants were not required to answer the questions to proceed to the next scenario. Upon completing the scenarios, participants answered questions about their age, gender, and nationality. Finally, participants reported their skiing ability, avalanche training, years of backcountry skiing experience, average number of backcountry skiing days per season, and past exposure to avalanche incidents. However, due to a technical failure, these covariates were not measured at all locations and are therefore not considered in our analysis.

Data preparation and analysis: We used R (R Core Team, 2017) for all data preparation and analyses. We reverse coded the risk judgment scores of participants in the Danger Group to make them comparable to the risk judgments of participants in the Safe Group. We henceforth refer to the judged risk as *judged safety* for both experimental conditions. There were missing values of judged safety (Study 1 = 2.3%; Study 2 = 5.1%; Study 3 = 1.0%; Study 4 = .5%; Study 5 = .1%; Study 6 = .6%) and behavior intention (Study 1 = 2.4%; Study 2 = 4.6%; Study 3 = 1.2%; Study 4 = .4%; Study 5 = .1%; Study 6 = 0%) within the data. We deemed those values to be missing at random. Mixed model analyses can handle missing values without requiring the exclusion of participants for whom only partial data was collected (Baayen et al., 2008). We therefore did not impute any data for missing values. For Studies 5 and 6, we subset the data by scenario category. We treated judged safety and behavior intention as ordinal variables. To examine whether the experimental condition influenced the odds of each value of judged safety, we used the *ordinal* package (Christensen, 2019) to implement cumulative link mixed models via Laplace approximations for the hypothesized model with judged safety as the outcome variable and the experimental condition as the predictor variable. We included intercepts for participants and scenarios as random effects to account for by-subject and by-scenario variability. We calculated Chi-square values (χ^2) with likelihood-ratio tests comparing the model that included the predictor variable(s) under investigation as the fixed effect (and participants and scenarios as random effects) against an equivalent model that excluded that predictor variable(s). To examine predictors of behavior, we estimated ordinal mixed models via Laplace approximations to analyze whether the predictor variables experimental condition, judged safety or their interaction predicted the odds of the outcome variable self-reported behavior intention. We defined the response order of behavior intention as "*No*" < "*I cannot say*" < "*Yes*". We treated the predictor judged safety as an interval variable for all analyses involving behavior intention as the outcome variable. We included intercepts for participants and scenarios as random effects, thereby accounting for by-subject and by-

scenario variability. We calculated Chi-square values (χ^2) with likelihood-ratio tests using the method described above.

Results

Studies 1 to 4: Single Reference Judgments of Uncertain Scenarios

Table 3 presents the proportions of judged safety values per experimental condition for Studies 1 to 4, with mode values clearly marked. The experimental condition influenced judged safety in all studies: Study 1, $\chi^2(1) = 30.49$, $p < .001$; Study 2, $\chi^2(1) = 12.25$, $p < .001$; Study 3, $\chi^2(1) = 19.65$, $p < .001$; and Study 4, $\chi^2(1) = 8.14$, $p = .004$. The log-odds coefficients and odds ratios for the main effect experimental condition for each study are presented in Table 4. Participants in the Danger Group who judged "*How dangerous is it?*" were at least 1.6 times more likely to judge a scenario to be safer than participants in the Safe Group who judged "*How safe is it?*". The probabilities of each value of judged safety per experimental condition are presented in Figure 1. In all four studies, judged safety was higher when judged in terms of how dangerous the scenario was than when judged in terms of how safe the scenario was.

Table 3 presents the proportions of behavior intention values per experimental condition for Studies 1 to 4, with mode response clearly marked. Judged safety predicted behavior intention in all four studies: Study 1, $\chi^2(1) = 1995.60$, $p < .001$, Study 2, $\chi^2(1) = 451.93$, $p < .001$, Study 3, $\chi^2(1) = 609.66$, $p < .001$, and Study 4, $\chi^2(1) = 388.44$, $p < .001$. The experimental condition alone, as a main effect, marginally predicted behavior in Study 2, $b = -.37$, $SE = .20$, $\chi^2(1) = 3.56$, $p = .059$, although that effect is mediated by the main effect judged safety (see Table 5). Otherwise, the experimental condition alone did not predict behavior in Studies 1, 3 or 4. The addition of the main effect experimental condition to the model with the main effect judged safety predicted behavior intention in Study 1, $\chi^2(1) = 17.76$, $p < .001$, and Study 3, $\chi^2(1) = 11.51$, $p < .001$, and marginally predicted behavior in Study 4, $\chi^2(1) = 2.69$, $p = .101$. The log-odds coefficients and odds ratios for the influence of the main effects judged safety and experimental condition on behavior intentions are presented in Table 5. Figure 2 illustrates the probability of each behavior per judged safety value per experimental condition. An interaction between the experimental condition and judged safety did not predict behavior intention in any of the four studies: Study 1, $\chi^2(1) = 1.56$, $p = .211$, Study 2, $\chi^2(1) = 1.05$, $p = .305$, Study 3, $\chi^2(1) = .02$, $p = .896$, or Study 4, $\chi^2(1) = .00$, $p = .949$. Overall we found that as the value of judged safety increased, the probability that participants in both experimental groups would ski the slope in the scenario increased. If participants in both experimental conditions judged safety to be equal, the probability that participants in the Safe Group would ski a slope was higher than that of participants in the Danger Group. However, based on the

magnitude of the log-odds coefficients and the odds ratios, judged safety had the greatest predictive power of behavior intention. Consequently, participants in the Safe Group were qualitatively less inclined to indicate that they would ski a slope because they were more likely to judge safety as lower.

To further investigate the robustness and scope of these measured effects, we conducted two additional studies using a broader set of scenarios with different levels of objective risk. Participants in Studies 1 to 4 judged scenarios that were all of uncertain risk. Is it possible that the high degree of uncertainty in some way influenced or accounted for the observed effect? Is uncertainty a prerequisite for the effect or does the question framing effect extend to situations that are objectively safe or dangerous? These are important questions because in a real world context, people encounter a range of situations of different objective risk with varying degrees of uncertainty. Understanding the contexts to which the observed question framing effect extends will inform strategies for its potential application. We therefore conducted Studies 5 and 6 using scenarios in three categories of objective risk—uncertain scenarios, safe scenarios, and dangerous scenarios—to examine if the effect observed in Studies 1 to 4 is present under varying degrees of objective risk and uncertainty.

Table 3

Percentage of judged safety values (values 1 to 7) and percentage of behavior (No = No, I would not ski the slope, UD = Undecided, I cannot say, and Yes = Yes, I would ski the slope) per experimental condition for Studies 1-4. The mode judged safety value and the mode behavior per experimental condition is marked by bold font.

Frame	Judged safety score							Behavior		
	1	2	3	4	5	6	7	No	UD	Yes
Study 1										
Safe	17.7	21.2	**22.0**	17.4	14.2	5.7	1.8	**49.7**	7.9	42.4
Danger	12.7	15.5	**23.1**	22.5	17.9	7.3	1.0	**49.9**	5.5	44.6
Study 2										
Safe	24.0	**27.0**	20.6	13.5	9.9	4.1	.9	**56.6**	8.9	34.5
Danger	18.9	17.5	**22.0**	21.0	14.1	6.1	.4	**52.2**	4.2	43.6
Study 3										
Safe	**22.9**	19.8	21.6	20.1	12.0	2.4	1.2	**47.1**	6.1	46.8
Danger	11.8	17.0	24.0	**25.3**	17.5	3.7	.7	45.9	7.7	**46.4**
Study 4										
Safe	25.2	**26.6**	24.1	16.1	5.6	2.0	.4	**61.8**	6.6	31.6
Danger	18.5	23.0	**23.4**	19.7	12.7	2.5	.2	**60.8**	3.5	35.7

Table 4

Coefficients of the model (judged safety ~ experimental condition) predicting whether the experimental condition influences judged safety for Studies 1-4

	$b_{\text{Danger frame}}$ (SE)	95% CI for Odds Ratios (OR)			p-value
		Lower	OR	Upper	
Study 1	.47 (.08)	1.36	1.60	1.89	< .001
Study 2	.63 (.18)	1.33	1.89	2.68	< .001
Study 3	.72 (.16)	1.50	2.04	2.79	< .001
Study 4	.60 (.21)	1.21	1.81	2.72	.004

Figure 1

Probabilities of judged safety values by experimental condition, with 95% confidence intervals, for Studies 1-4.

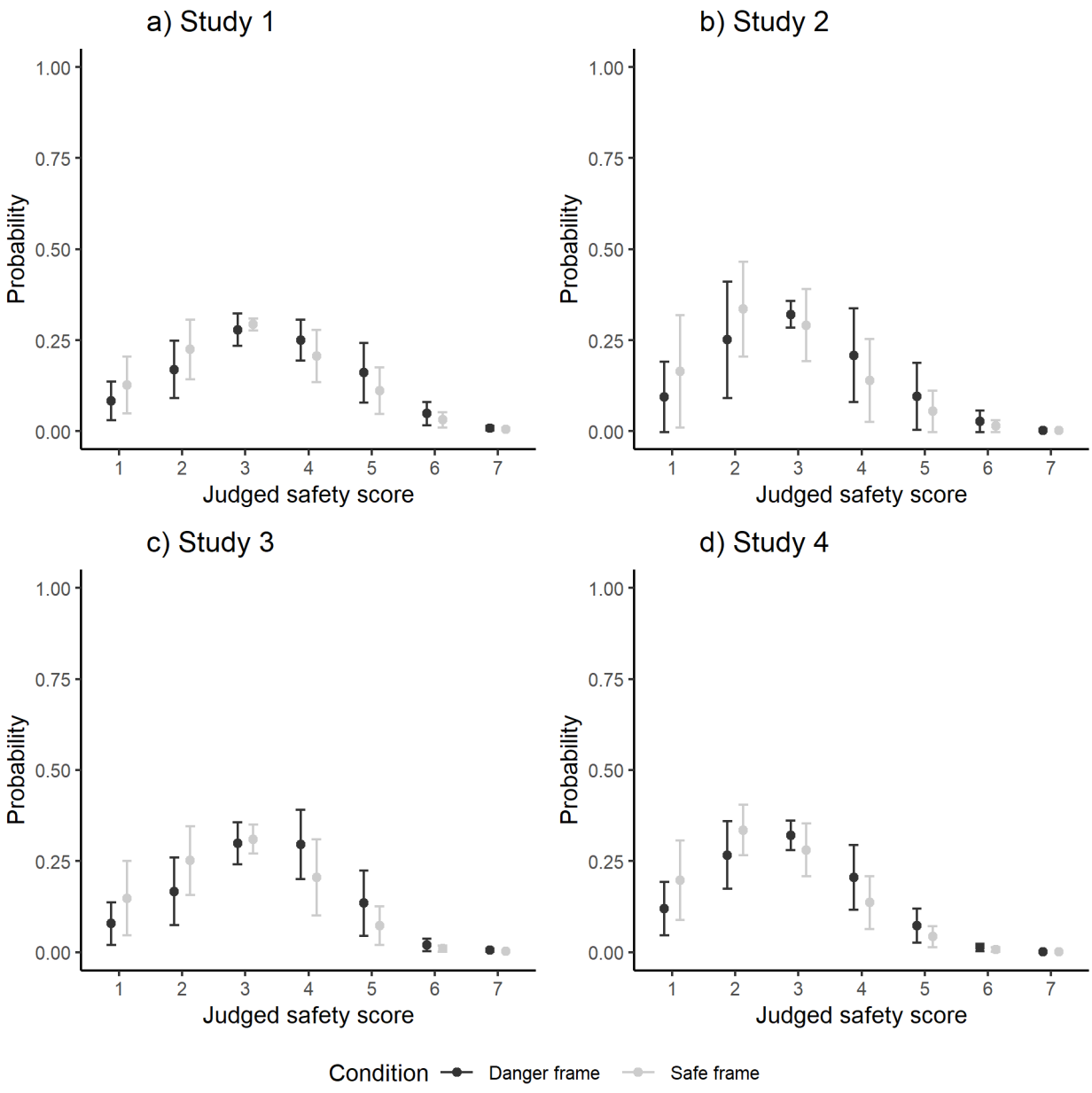


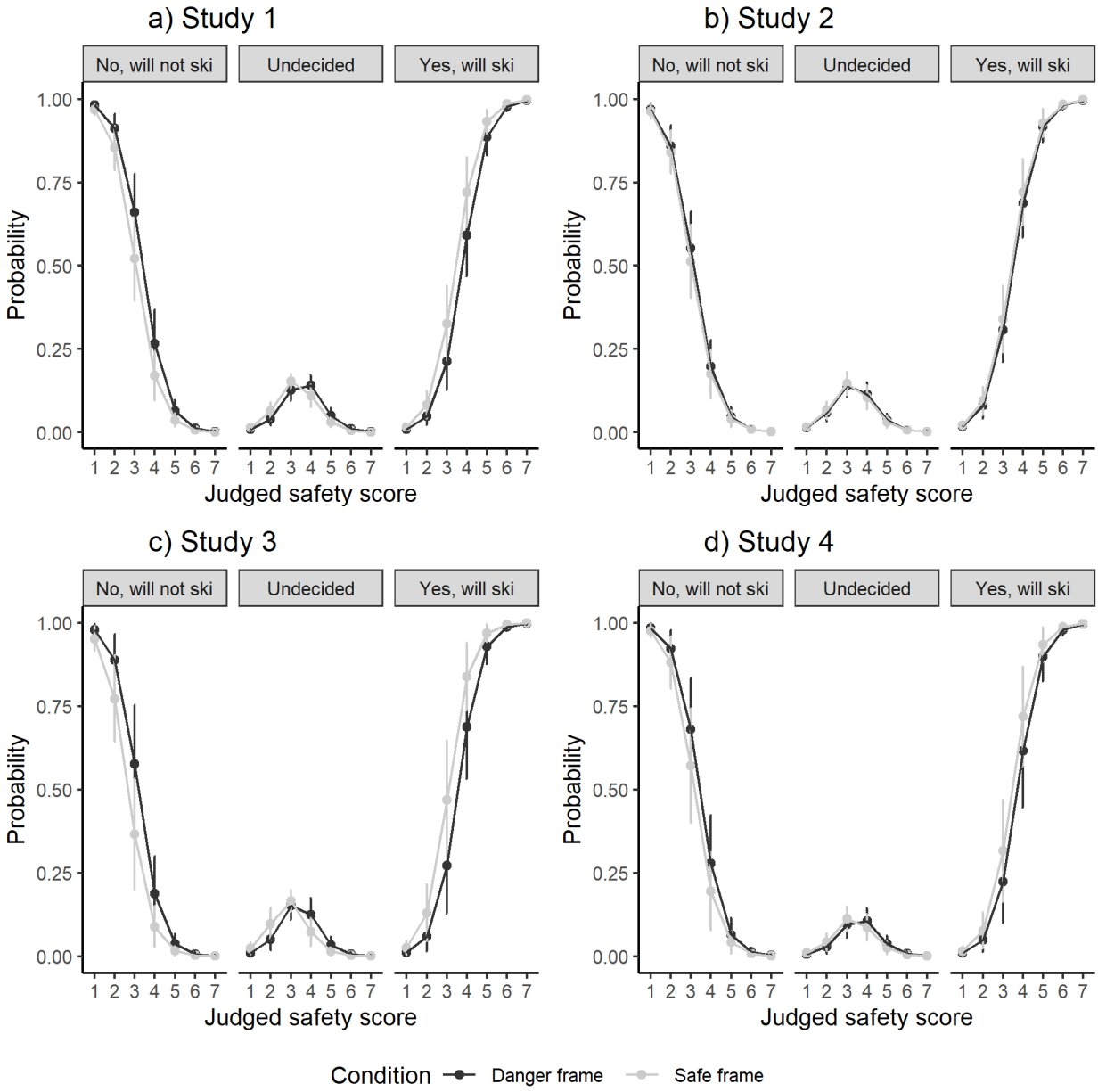
Table 5

Coefficients of the model (behavior ~ judged safety + experimental condition) predicting whether the terms judged safety and experimental condition influences behavior intention in Studies 1-4

	b (SE)	95% CI for Odds Ratios (OR)			p-value
		Lower	OR	Upper	
Study 1					
Judged safety	1.67 (.06)	4.76	5.34	5.98	< .001
Safe frame	.57 (.14)	1.36	1.78	2.32	< .001
Study 2					
Judged safety	1.60 (.11)	3.98	4.98	6.23	< .001
Safe frame	.15 (.26)	.69	1.17	1.96	.562
Study 3					
Judged safety	1.77 (.11)	4.78	5.90	7.29	< .001
Safe frame	.86 (.25)	1.43	2.36	3.88	< .001
Study 4					
Judged safety	1.71 (.13)	4.30	5.54	7.13	< .001
Safe frame	.47 (.29)	.91	1.60	2.80	.101

Figure 2

Probabilities, with 95% confidence intervals, of skiing the scenario slope by judged safety values and experimental condition for Studies 1-4.



Studies 5 and 6: Single Reference Judgments of Uncertain, Safe and Dangerous Scenarios

Table 6 presents the proportions of judged safety values per experimental condition for each category of scenarios in Studies 5 and 6, with the mode values clearly marked. In Study 5, the experimental condition influenced judged safety—presented according to the order in which participants judged the categories of scenarios—for dangerous scenarios, $\chi^2(1) = 52.59$, $p < .001$, and scenarios of uncertain risk, $\chi^2(1) = 10.28$, $p = .001$, but did not influence the judged safety for safe scenarios, $\chi^2(1) = .55$, $p = .457$, which participants judged last. In Study 6, the experimental condition influenced judged safety—presented according to the order in which participants judged the categories of scenarios—of safe scenarios, $\chi^2(1) = 12.45$, $p < .001$, did not influence the judged safety of scenarios of uncertain safety, $\chi^2(1) = 2.29$, $p = .130$, and influenced the judged safety of dangerous scenarios, $\chi^2(1) = 85.37$, $p < .001$. The log-odds coefficients and odds ratios for the main effect experimental condition per scenario category are presented in Table 7. The probabilities of each value of judged safety per scenario category by experimental condition are presented in Figure 3. Apart from the safe scenarios that were judged last in Study 5 and the uncertain scenarios that were judged second in Study 6 for which there was no effect, participants in the Danger Group were more likely to judge all categories of scenarios to be safer than participants in the Safe Group. Study 5 and Study 6 replicated the result that framing the risk judgment in terms of safety was more likely to result in lower judged safety whereas framing the risk judgment in terms of danger was more likely to result in higher judged safety.

Table 6 presents the proportions of behavior intention values per experimental condition for Studies 5 and 6, with the mode response clearly marked. The main effect judged safety predicted behavior intention for safe scenarios in Study 5, $\chi^2(1) = 112.21$, $p < .001$, and Study 6, $\chi^2(1) = 113.67$, $p < .001$; for uncertain scenarios in Study 5, $\chi^2(1) = 322.23$, $p < .001$, and Study 6, $\chi^2(1) = 133.74$, $p < .001$; and for dangerous scenarios in Study 5, $\chi^2(1) = 28.32$, $p < .001$, and Study 6, $\chi^2(1) = 3.14$, $p < .076$. The experimental condition alone, as a main effect, predicted behavior for safe scenarios in Study 6, $\chi^2(1) = 8.47$, $p < .003$; an effect that is mediated by the main effect judged safety (see Table 8). Otherwise, the experimental condition alone did not predict behavior intention in either Study 5 or Study 6. The addition of the main effect experimental condition to the model with the main effect judged safety predicted behavior intention for uncertain scenarios in Study 5, $\chi^2(2) = 6.41$, $p = .011$, and an interaction between judged safety and the experimental condition predicted behavior for the safe scenarios in Study 5, $\chi^2(1) = 6.78$, $p = .009$. However, the experimental condition had no influence on behavior for the safe scenarios or the uncertain scenarios in Study 6 beyond what was predicted by judged

safety. The log-odds coefficients and odds ratios for the influence of the main effect judged safety together with the main effect experimental condition (and their interaction, if relevant) on behavior intentions are presented in Table 8. As the value of judged safety increased, the probability that participants in both experimental groups would ski the slope increased for safe scenarios and scenarios of uncertain risk. Studies 5 and 6 replicated the result that judged safety had the greatest predictive power of behavior intention. Participants in the Safe Group were therefore qualitatively less inclined to indicate that they would ski a slope because they were more likely to judge safety as lower. As for dangerous scenarios, there were too few measures of judged safety values greater than 4 in Study 5 and greater than 3 in Study 6 to reliably test for an interaction effect between judged safety and experimental condition. This is indicated by the range of the 95% CI in Figure 4. Participants in both experimental groups in Studies 5 and 6 were effectively unanimous that it was 100% likely that they would not ski the slopes in the dangerous scenarios.

Table 6

Percentage of judged safety values (values 1 to 7) and percentage of behavior (No = No, I would not ski the slope, UD = Undecided, I cannot say, and Yes = Yes, I would ski the slope) per experimental condition (column Group) and per scenario category (column Scenario, S = safe, UR = uncertain risk, and AD = dangerous) for Studies 5 and 6. The mode judged safety value and the mode behavior per experimental condition and per scenario category is marked by bold font.

Frame	Scenario	Judged safety score							Behavior		
		1	2	3	4	5	6	7	No	UD	Yes
Study 5											
Safe	Safe	1.0	3.1	6.6	11.7	22.3	**34.5**	20.8	8.6	2.0	**89.4**
Danger	Safe	.0	1.0	6.8	10.8	23.9	**37.9**	19.6	5.6	2.9	**91.5**
Safe	Uncertain	11.7	27.6	**29.6**	19.7	9.8	.8	.8	**59.1**	9.1	31.8
Danger	Uncertain	8.6	15.9	26.5	**30.4**	14.7	3.4	.5	**57.1**	7.6	35.3
Safe	Dangerous	**66.2**	25.8	6.0	2.0	.0	.0	.0	**96.0**	2.5	1.5
Danger	Dangerous	**37.2**	24.5	15.7	13.4	6.9	1.6	.7	**95.1**	.3	4.6
Study 6											
Safe	Safe	3.2	7.4	14.9	18.1	**28.7**	14.9	12.8	16.7	13.5	**69.8**
Danger	Safe	.8	2.6	6.8	9.4	29.1	**29.9**	21.4	10.2	2.6	**87.2**
Safe	Uncertain	26.0	**32.3**	23.6	9.4	5.5	2.4	.8	**68.7**	13.3	18.0
Danger	Uncertain	21.9	22.0	**31.6**	14.2	10.3	.0	.0	**69.2**	10.3	20.5
Safe	Dangerous	**92.7**	6.3	1.0	.0	.0	.0	.0	**96.9**	2.1	1.0
Danger	Dangerous	**55.6**	13.7	20.5	6.0	4.3	.0	.0	**97.4**	2.6	.0

Table 7

Coefficients of the model (judged safety ~ experimental condition) predicting whether the experimental condition influences judged safety per scenario category for Studies 5 and 6.

	$b_{\text{Danger frame}}$ (SE)	95% CI for Odds Ratios (OR)			p-value
		Lower	OR	Upper	
Study 5					
Safe scenarios	.20 (.27)	.72	1.22	2.08	.457
Uncertain scenarios	.88 (.27)	1.42	2.42	4.15	.001
Dangerous scenarios	1.97 (.28)	4.14	7.17	12.42	< .001
Study 6					
Safe scenarios	1.62 (.46)	2.07	5.08	12.45	< .001
Uncertain scenarios	.43 (.28)	.88	1.53	2.66	.130
Dangerous scenarios	3.35 (.64)	8.06	28.42	100.18	< .001

Figure 3

Probabilities of judged safety values per experimental condition and per scenario category, with 95% confidence intervals, for Studies 5 and 6.

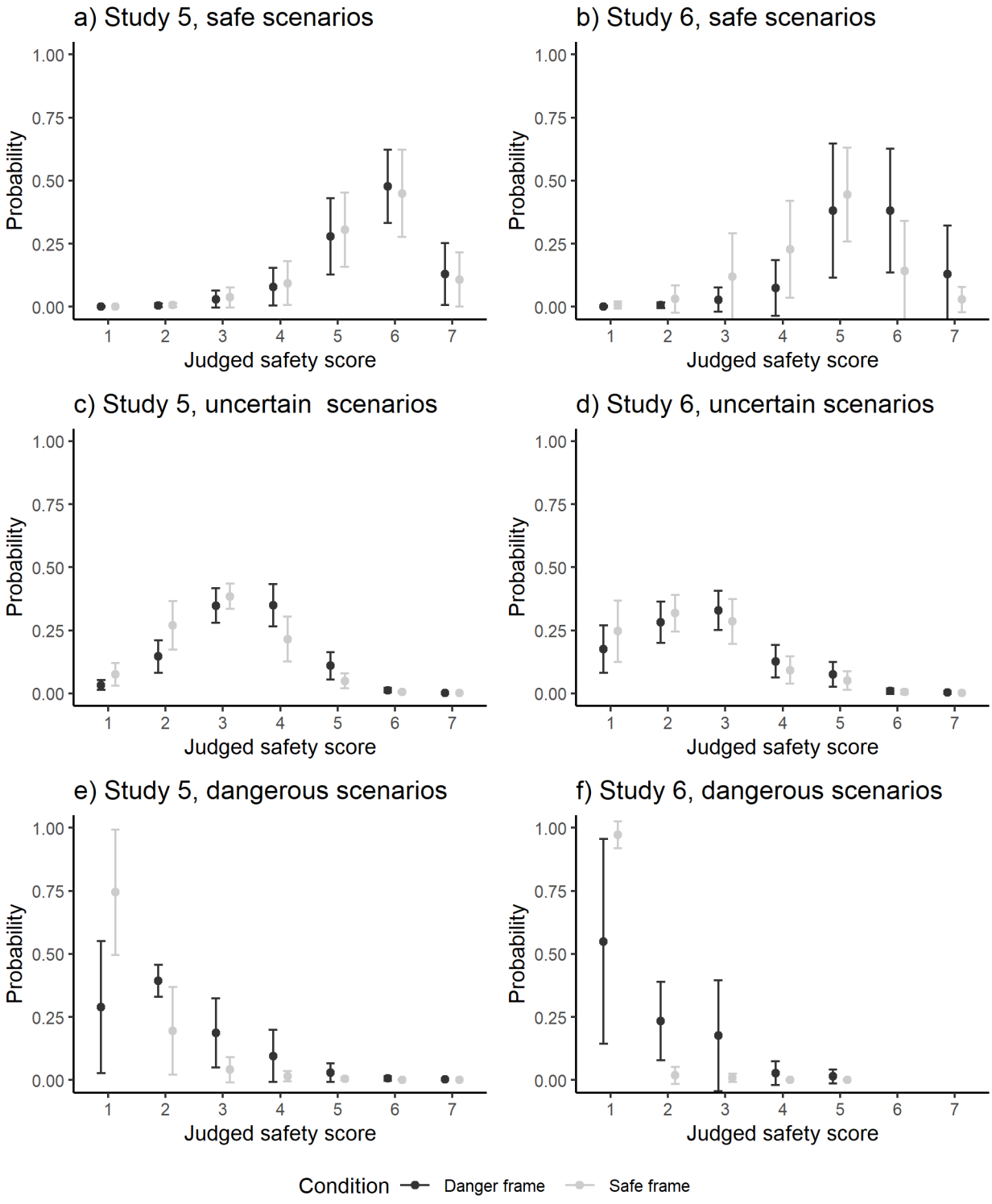


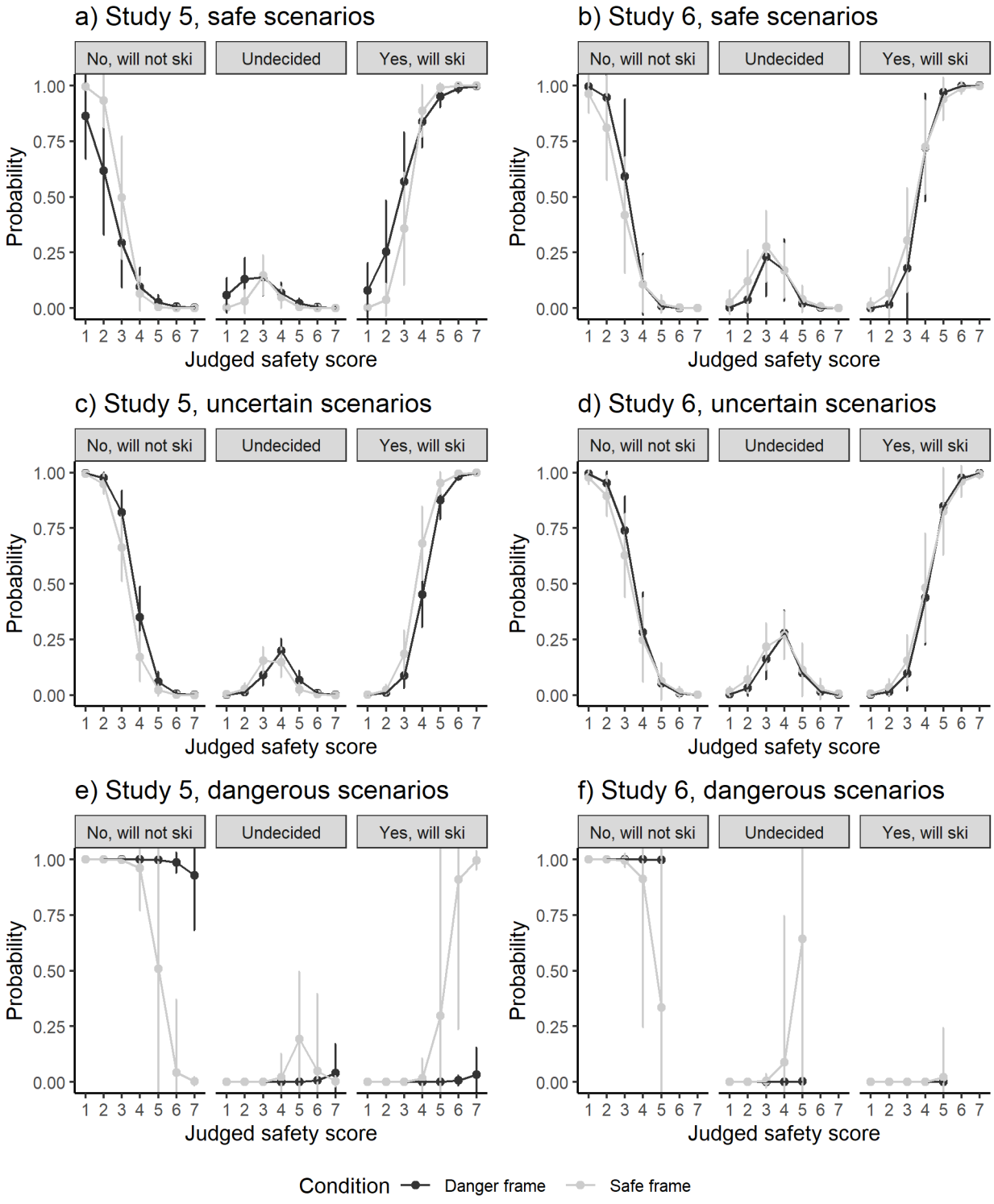
Table 8

Coefficients of the model predicting whether the terms judged safety and experimental condition influences behavior intention per scenario category in Studies 5 and 6

	b (SE)	Odds Ratios (OR) 95% CI			p-value
		Lower	OR	Upper	
Study 5, safe scenarios					
Judged safety	1.36 (.28)	2.25	3.90	6.75	< .001
Safe frame	-4.71 (2.23)	.00	.00	.71	.034
Judged safety * safe frame	1.28 (.59)	1.14	3.60	11.39	.029
Study 5, uncertain scenarios					
Judged safety	2.18 (.19)	6.11	8.88	12.91	< .001
Safe frame	.89 (.35)	1.22	2.44	4.88	.012
Study 6, safe scenarios					
Judged safety	2.02 (.46)	3.09	7.56	18.46	< .001
Safe frame	.11 (.55)	.37	1.11	3.33	.849
Study 6, uncertain scenarios					
Judged safety	1.79 (.25)	3.68	6.00	9.78	< .001
Safe frame	.50 (.47)	.65	1.65	4.19	.288

Figure 4

Probabilities, with 95% confidence intervals, of skiing the scenario slope by judged safety values per experimental condition per scenario category for Studies 5 and 6.



General Discussion

Results from six studies demonstrate that risk perception is influenced by the frame of the question that elicits the risk judgment. Framing a risk judgment in terms of safety ("*How safe is it?*") was more likely to result in lower judged safety than framing that risk judgment in terms of danger ("*How dangerous is it?*"). The results of Studies 5 and 6 indicate that uncertainty is not a prerequisite for this framing effect. The question frame had a consistent direction of effect when judging safe scenarios, uncertain scenarios, and dangerous scenarios, suggesting that, in principle, the effect from framing risk judgments in terms of safety or danger applies in all situations of objective risk.

Question Frames Guide Attention during Information Sampling

A question frame elicits a judgment that is relative to the reference defined by the frame. This process can be thought of as analogous to testing the hypothesis defined by the frame. The question "*How safe is it?*" defines *safe* as the provisional reference point for the risk judgment, whereas the question "*How dangerous is it?*" defines dangerous as the provisional reference point for the risk judgment. This has the effect of focusing the decision maker's attention on selectively sampling evidence to evaluate whether the target of judgment meets or fulfills the descriptive state defined by the question frame rather than judging the degree of risk according to the two complementary poles *completely safe* and *completely dangerous*. Participants presented the safe frame judged whether the available evidence was sufficient to establish whether a scenario was indeed safe. Participants presented the danger frame judged the available evidence with a focus on ascertaining whether the same scenario was indeed dangerous. To ask how safe a scenario is or to ask how dangerous it is are therefore not informationally equivalent frames, despite being complementary dimensions of the bipolar attribute *risk*. Each question focuses the decision maker's attention on selectively sampling different evidence in relation to different reference points (Chong & Druckman, 2007; Druckman, 2001; Keren, 2011; Teigen, 2015), effectively making each question a different judgment task. This is particularly relevant in a decision environment such as avalanche terrain where information cues rarely have logically equivalent opposites. There is an asymmetry of relevant evidence between the frames. Although the presence of one sign indicates increased danger and the greater probability of an avalanche, that sign may have no logically equivalent opposite. Moreover, the absence of that sign is not necessarily an indication of increased safety.

Participants in our studies sampled different evidence in relation to the reference point defined by the question frame and reported their judgment on a scale similarly defined by that reference point. As illustrated in Figure 5, participants searched for evidence of safety if *safe*

framed the question eliciting the risk judgment, or searched for evidence of danger if *danger* framed the question. Yet under conditions of high uncertainty, there was insufficient evidence to definitively establish or reject the descriptive state that either reference point emphasized. The evidence in favor of either reference point was ambiguous and therefore participants judging safety concluded that a scenario was not definitively safe, whereas participants judging danger similarly concluded that the same scenario was not entirely dangerous. Under both frames, participants adjusted their reported judgment according to the perceived (in)sufficiency of evidence for their respective reference point defined by the question frame. Those adjustments were made in relation to the limits, also defined by the question frame, of their respective reporting scales. However, the available evidence and consequently the adjustment on the response scale differed between the framing conditions. As a result, participants who were presented the safe frame judged the scenarios to be relatively more dangerous (alternatively, relatively less safe), while participants who were presented the danger frame judged the same scenarios to be relatively safer (alternatively, relatively less dangerous).

Interestingly, we observed this effect when participants judged dangerous and safe scenarios. Despite the availability of what avalanche experts consider to be definitive evidence of the objective safety and danger of the scenarios, participants who faced the safe frame did not accept the reference point *safe* for safe scenarios, and consequently judged them to be relatively less safe than participants who faced the danger frame did. Similarly, when judging dangerous scenarios, those who faced the danger frame did not accept the reference *danger* defined by that frame, and consequently judged dangerous scenarios to be relatively less dangerous (i.e., more safe) than participants under the safe frame. Objective uncertainty was not a requirement for the question frame to evoke selective evidence sampling relative to the reference point defined by the frame. Although a framing effect was not measured for judgments of safe scenarios in Study 5 and uncertain scenarios in Study 6, we suspect this to be the result of the order in which participants judged the scenarios. When judging scenarios of different objective risk levels, participants might have judged a given scenario in relation to the sufficiency of evidence and the judged safety/danger of the previous scenario(s). We also suspect that an order effect was the cause of the different magnitude of measured effects between scenarios categories in Studies 5 and 6. Examining the role of presentation order more directly may be an interesting direction for future work, but a challenge would be to ensure that such an investigation is ecologically valid. It is unlikely that a person would judge vastly different conditions in close temporal proximity. Thus the order effects we observed here may have been, to some extent, an artifact of overly stark contrasts between objective risk levels.

The Possibility of Other Cognitive Processes during Question Frames

The direction of the question framing effect we observed across six studies runs counter to what might be expected based on alternative accounts of framing that assume a valence or communication driven mechanism. In particular, both the valence account and the communication account of framing would be consistent with judged safety being higher when judging "How safe is it?" Earlier research found on the direction of the question framing effect and the underlying cognitive processes found conflicting results. Payne and colleagues (2013) found credible evidence of a question framing effect consistent with the valence account of framing. Although Comerford and Robinson (2017) replicated the results of Payne and colleagues when testing the same judgment task, when they replaced the reporting format from a probabilistic estimate to a point estimate, their results indicated a framing effect in the opposite direction.

This apparent mismatch between the predictions of other framing accounts and our findings may be linked to the specific framing paradigm we used. In addition to specifying a verbal question frame (how safe vs. dangerous is it?), the judgment task we used involved sampling information beyond what was described by the question frame. This process of information sampling is not a common element in most judgment tasks used in other research on framing, and the cognitive processes underlying framing effects likely depend on the method by which the frame is achieved. Kreiner and Gamliel (2018), for instance, found evidence that attention mechanisms contribute to attribute framing but recognized that their experimental design prevented them from ruling out the potential contribution of valence mechanisms to the overall framing effect. Similarly, in our paradigm, attentional mechanisms may have played a dominant role because we included an information-sampling component and because the frame was achieved by the phrasing of a question (rather than a statement). Yet we cannot rule out that valence-driven and communicative mechanisms also played a role in the results of our study. Participants' initial prior when prompted with the question *How safe is it?* may well have leaned toward the valence of that frame, but was then revised in the process of gathering insufficient evidence to support this initial hypothesis. Thus, framing may be an aggregation of different forces resulting from distinct cognitive processes that are evoked by the judgment task and the manner in which the frame is achieved. An interesting avenue for future research is to use computational cognitive modeling to disentangle the cognitive processes that drive question framing.

Figure 5

A conceptualization of the decision making processes under each experimental framing condition. The person on the left represents the Safe Group judging risk in response to the question "How safe is it?" The person on the right represents the Danger Group judging risk in response to the question "How dangerous is it?" The thought bubbles illustrate the assumed cognitive process up to the behavior decisions expressed in the speech bubbles. The icons in the top left and the "35°"are examples of the information provided in the scenarios. They represent the forecasted regional avalanche danger rating (3), the avalanche problem, and the slope angle.

The Indirect Influence of Question Framing on Behavior Decisions

Beyond the effect of question framing on judged safety, the results from all six studies indicate that judged safety influenced participants' hypothetical behavior intention. The probability that participants would ski the slope monotonically increased with an increase in judged safety. The only exception to this clear result concerned the judgments of dangerous scenarios, for which participants were effectively unanimous that irrespective of judged safety they would not ski the slope (see Figure 4). However, it is very rare for dangerous slopes to be so clearly marked as dangerous, as they were in our studies, by evidence of an active avalanche on that particular slope. Typical of a wicked learning environment (Hogarth et al., 2015), a dangerous slope commonly resembles an uncertain slope until someone travels on it triggering an avalanche, thus providing clear but rare evidence of the objective risk level, albeit a little too late.

Indecision was a response option for the measure of behavior intention, and a small proportion of responses (ranging from .3 to 13.5% across all studies and scenario categories) indicated that participants were undecided about their intended action. The likelihood of such indecision was highest when scenarios were perceived to be neither completely safe nor completely dangerous, with judged safety values in the middle range of the response scale. This establishes that the judged safety response scale captured the equivalent poles of the bipolar attribute of judgment, despite only one of those poles framing the judgment task. However, it is important to point out that indecision is not a true response option in a real-world situation; either skiing the slope or not skiing the slope are only ever observed.

Behavior is what exposes people to risk. That judged safety influenced behavior intention is particularly important for our examination of question framing and the possibility of harnessing that framing effect to promote safer risk judgments and decisions. Although there was no overall effect of framing on hypothetical behavioral intent, it is important to emphasize the process by which question framing influenced behavior. Question framing was found to influence judged safety, and judged safety was in turn found to be a powerful determinant of behavior decisions. Since behavior is the immediate cause of risk exposure, any factor that can directly or indirectly influence behavior can potentially be utilized to reduce the frequency of accidents and fatalities. Phrasing the risk judgment in terms of how safe the slope is resulted in lower values of judged safety, which in turn resulted in a lower likelihood of deciding to ski the slope. This suggests that judging how safe a risk is will result in the safest behavior with respect to the potential risk. The safe frame was found to indirectly result in more cautious behavior via the direct effect of framing on judged safety. We illustrate the indirect influence of question

framing on behavior intention via the effect on judged safety in Figure 5. This illustration is intended to make clear both the presumed cognitive processes and the potential applied relevance of question framing to the widest possible audience. By selectively phrasing the question that elicits a risk judgment, a decision maker's attention can be directed in a way that strategically influences the perception of risk with the effect of making one behavior outcome more likely.

Interestingly, at any given level of judged safety—if participants in the two experimental groups judged safety to be the same—those prompted with the safe frame indicated that they were more likely to ski the slope than were participants prompted with the danger frame. To understand this apparently contradictory effect, consider the conditions under which judged safety will be equal between the two framing conditions. Due to the question framing effect, the judged safety of a given scenario was more likely lower under the safe frame and higher under the danger frame. That framing effect must be offset or overcome in order for judged safety between the two framing conditions to be equal. We can therefore assume that when judged safety under the two frames was equal, the perceived evidence basis for the judgments were not equivalent. The intention to ski may have been higher under the safe frame because that decision maker perceived more evidence of safety (more in terms of validity, relevance, weight, or even volume of evidence), and/or the intention to ski may be lower under the danger frame because that decision maker perceived similarly more evidence of danger. The behavior decisions under each frame are based on an asymmetry of evidence, an asymmetry that was necessary to offset the framing effect in order for judged safety to be equal. Although we appear to find more risk acceptance under the safe frame when judged safety between the experimental conditions is equal, the behavior decisions under the safe frame are potentially made on a more valid, relevant sample of evidence. Importantly, however, this finding must be placed in the context that judged safety was the strongest predictor of behavior intention. Participants who judged how safe a scenario is were more likely to judge safety as lower, and the likelihood of skiing a slope decreased as judged safety decreased. The safe frame was found to indirectly result in more cautious behavior via the direct effect of framing on judged safety.

Implications for Applied Risk Judgments and Risk Communication

Backcountry skiing in avalanche terrain exemplifies a crucial challenge in applied risk communication and risk perception: people desire to engage in an activity despite knowing the inherent risk of serious injury or death. Avalanche accidents are overwhelmingly the result of human error. In 90% of fatal avalanche accidents, the victim or someone in the victim's party triggered the avalanche (McClung & Schaerer, 2006), implying that people's risk perception

and decisions are critical factors in avalanche fatalities. Information on the conditions in avalanche terrain such as the complex conditions of the snowpack, its metamorphosis over time, and the effects of terrain and weather, together with the knowledge of how to use this information are essential for judging avalanche risk. In an attempt to reduce the number of accidents and fatalities, stakeholders such as national avalanche warning services and education providers have done much work to provide detailed avalanche forecasts and improve public knowledge of the dangers and best practices for safety in avalanche terrain (for a review, see e.g., Engeset et al, 2018). Yet despite these efforts, avalanches continue to claim the lives of a troubling number of participants in this increasingly popular and unregulated activity. The dissemination of information—which as a stand-alone activity is a failed strategy for changing how people perceive risk and behave (Kelly & Barker, 2016; Simis et al, 2016)—has been insufficient for ensuring avalanche safety among backcountry skiers. Might question framing serve as a complementary strategy to promote more cautious risk perception and behavior?

Decision makers, avalanche warning services, and education providers have substantial control over the formulation of questions about the risks assessed during a backcountry ski tour. Our findings illustrate how the language used to formulate risk judgments and its influence on the cognitive processes has the potential for a real and tangible impact on how people perceive risk and, ultimately, behave in the face of risk. Selectively framing risk judgments might serve as one effective component of a multifaceted strategy to promote more cautious and conservative decisions in avalanche terrain and other domains involving risks. These findings have potential real-world application in teaching methods, tools, and strategies for reducing accidents and fatalities. At the public and institutional level such as a national or regional avalanche forecasting service, the frame used when presenting information about conditions in avalanche terrain might influence how users (i.e. the general public) both perceive the current risk and, more critically, how they decide to act. At the individual or group level, communication between members of a group travelling in avalanche terrain, while typically presented with little thought toward framing, could be positively impacted by increased awareness of the framing effect. Specifically, if a group member is presenting route options or tour alternatives, the way in which information and questions are framed could influence other group members' perceptions of the current risk and the decisions that are made or communicated by members of the group.

Despite the apparent promise, there are several important considerations and potential limitations for the application of question framing to increase skier safety in avalanche terrain, or to promote certain judgments and decisions within any other discipline or context. It is

currently an open question whether people can prompt themselves to frame questions about risky situations in a way that promotes safer judgments, highlighting an interesting avenue for future research. It is unlikely that how a decision maker internally represents the problem or judgment is entirely determined by externally presented information and/or the formulation of the judgment task (Tversky & Kahneman, 1981). Decision makers clearly use their own experience and knowledge when modelling the world in order to judge the probabilities of potential outcomes (e.g., Wulff et al., 2019), and they may automatically do so with a familiar, default reference when not prompted with a question frame. Another important consideration is whether *actual behavior* in the mountains, when judging a slope to really ski while facing the real risk of avalanche, would be affected differently than hypothetical behavioral intent measured in an auditorium using fictional scenarios. Behavior often deviates from intention (Sheeran & Webb, 2016). Moreover, contextual cues only encountered in the natural decision environment can reduce an anticipated framing effect (Bless et al., 1998). A third consideration is that we may fail to see the same framing effect outside of the experimental setting. Unlike a natural setting, participants in our study had limited time to judge the risk and decide their behavior intention for each scenario. Although there is conflicting evidence as to whether more thought reduces framing effects (for example, see LeBoeuf & Shafir, 2003; Martiny-Huenger et al, 2020), we may fail to see the same effect in a natural environment where decision makers have more time for reflection and where the potential consequence of error is considerably greater. A fourth consideration is the necessity to establish whether people's natural decision making process is to first judge risk in terms of safety or danger before then deciding behavior. The risk judgment itself may be a contrivance of the experimental design. Outside of an experimental setting, the behavior decision may encompass the risk judgment. Finally, any application of these findings should be tested in an applied setting before prescribing them for use in practice.

Research on framing effects informs policies and practices in other applied domains such as health (e.g., Garcia-Retamero & Cokely, 2011; Garcia-Retamero & Galesic, 2010; Peters et al., 2011; Rothman & Salovey, 1997) and finance (Kirchler et al, 2005; Weber et al., 2000). Our findings highlight a promising direction for the strategic application of question framing for increased safety in various domains of applied risk perception and communication. The aim of public risk management is to optimize the decision matrix to enable users to maximize their personal enjoyment and benefit while minimizing both individual and collective public risk. Critically, as in avalanche terrain, one wants to minimize the probability that a "*go*" decision is made under objectively "*no-go*" conditions. We do not believe that the framing of

risk judgment questions alone will be sufficient to ensure safe behavior among all decision makers and or in all risky domains. It is no substitute for the availability of valid evidence of the objective risk, and the necessary knowledge and experience to understand and apply that information. However, the adoption of a procedure strategy such as that afforded by framing risk judgment questions may boost an individual's overall competency for risk judgments or behavioral decisions. Various disciplines can conceivably harness the questions framing effect to make desired judgments and behaviors more likely. Risk management strategies, tools and education should recognize and account for this effect, and leverage these emergent findings to reduce the potential for accidents and fatalities.

Conclusion

The present research makes several contributions to the existing literature on framing effects. First, our research establishes the direction of the framing effect when a risk judgment is framed in terms of safety or danger. Risk judgments framed in terms of safety (*How safe is it?*) result in more cautious, conservative judgments than when framed in terms of danger (*How dangerous is it?*). Second, uncertainty was not a requirement for that effect. There was a framing effect when judging risk under varying degrees of uncertainty, be it under conditions of objective safety, uncertainty, or danger. These findings suggest that the question frame directed attention in a way that guided selective evidence sampling, rather than indicating a valence-consistent or communication-driven framing effect. Finally, our findings demonstrate the indirect influence of the question frame on behavior intention. The adoption of a procedure strategy such as that afforded by framing risk judgment questions can boost people's natural decision making competencies in order to ensure safer risk perceptions and behavior. These findings have the potential to inform the development of policies and practices that harness question framing in domains of applied risk perception and risk communication.

References

Almashat S, Ayotte B, Edelstein B, Margrett J. (2008). Framing effect debiasing in medical decision making. *Patient Education and Counseling, 71*(1):102-7. https://doi.org/10.1016/j.pec.2007.11.004

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412. https://doi.org/10.1016/j.jml.2007.12.005

Biswas, D. (2009), The effects of option framing on consumer choices: Making decisions in rational versus experiential processing modes. *Journal of Consumer Behaviour, 8*(5), 284-299. https://doi.org/10.1002/cb.288

Biswas, D. and Grau, S.L. (2008), Consumer choices under product option framing: Loss aversion principles or sensitivity to price differentials? *Psychology & Marketing, 25*, 399-415. https://doi.org/10.1002/mar.20217

Bless, H., Betsch, T. and Franzen, A. (1998), Framing the framing effect: the impact of context cues on solutions to the 'Asian disease' problem. *European Journal of Social Psychology, 28*. 287-291. https://doi.org/10.1002/(SICI)1099-0992(199803/04)28:2<287::AID-EJSP861>3.0.CO;2-U

Block, L., & Keller, P. (1995). When to Accentuate the Negative: The Effects of Perceived Efficacy and Message Framing on Intentions to Perform a Health-Related Behavior. J*ournal of Marketing Research, 32*(2), 192-203. https://doi.org/10.2307/3152047

Bruine de Bruin, W. (2011). Framing effects in surveys: How respondents make sense of the questions we ask. In G. Keren (Ed.), *Perspectives on Framing*. (pp. 3-33). New York, NY, US: Psychology Press.

Bruine de Bruin, W., van der Klaauw, W., Topa, G., Downs, J. S., Fischhoff, B., & Armantier, O. (2012). The effect of question wording on consumers' reported inflation expectations. *Journal of Economic Psychology, 33*(4), 749-757. https://doi.org/10.1016/j.joep.2012.02.001

Bui, T. C., Krieger, H. A., & Blumenthal-Barby, J. S. (2015). Framing Effects on Physicians' Judgment and Decision Making. *Psychological Reports, 117*(2), 508–522. https://doi.org/10.2466/13.PR0.117c20z0

Cacciatore, M. A., Scheufele, D. A., & Iyengar, S. (2016). The End of Framing as we Know it … and the Future of Media Effects. *Mass Communication and Society, 19*(1), 7-23. https://doi.org/10.1080/15205436.2015.1068811

Chong, D., & Druckman, J. N. (2007). Framing Theory. *Annual Review of Political Science, 10*(1), 103-126. doi:10.1146/annurev.polisci.10.072805.103054

Christensen, R. H. B. (2019). *ordinal - Regression Models for Ordinal Data* (Version 2019.4-25). http://www.cran.r-project.org/package=ordinal/

Comerford, D. A., and Robinson, J. (2017) Die-by Framing both Lengthens and Shortens Life: Further Evidence on Constructed Beliefs in Life Expectancy. *Journal of Behavioral Decision Making*, 30: 1104– 1112. https://doi.org/10.1002/bdm.2027

Druckman, J. N. (2001). Evaluating framing effects. *Journal of Economic Psychology, 22*(1), 91-101. https://doi.org/10.1016/S0167-4870(00)00032-5

Engeset, R. V., Pfuhl, G., Landrø, M., Mannberg, A., & Hetland, A. (2018). Communicating public avalanche warnings – what works? *Natural Hazards and Earth System Sciences, 18*(9), 2537-2559. https://doi.org/10.5194/nhess-18-2537-2018

Garcia-Retamero, R., & Cokely, E. T. (2011). Effective communication of risks to young adults: Using message framing and visual aids to increase condom use and STD screening. *Journal of Experimental Psychology: Applied, 17*(3), 270–287. https://doi.org/10.1037/a0023677

Garcia-Retamero, R., & Dhami, M. K. (2013). On avoiding framing effects in experienced decision makers. *Quarterly Journal of Experimental Psychology, 66*(4), 829-842. https://doi.org/10.1080%2F17470218.2012.727836

Garcia-Retamero, R., & Galesic, M. (2010). Who profits from visual aids: overcoming challenges in people's understanding of risks [corrected]. *Social Science & Medicine, 70*(7), 1019-1025. https://doi.org/10.1016/j.socscimed.2009.11.031

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics. Volume 3: Speech acts* (pp. 41–58). New York: Academic Press.

Hedger, N., Gray, K. L. H., Garner, M., & Adams, W. J. (2016). Are visual threats prioritized without awareness? A critical review and meta-analysis involving 3 behavioral paradigms and 2696 observers. *Psychological Bulletin, 142*(9), 934-968. http://dx.doi.org/10.1037/bul0000054

Hilton, D. J. (2011). Linguistic polarity, outcome framing, and the structure of decision making: A pragmatic approach. In G. Keren (Ed.), *Perspectives on Framing*. (pp. 135-155). New York, NY, US: Psychology Press.

Hogarth, R. M., Lejarraga, T., & Soyer, E. (2015). The two settings of kind and wicked learning environments. *Current Directions in Psychological Science, 24*(5), 379-385. https://doi.org/10.1177/0963721415591878

Hohle, S. M., & Teigen, K. H. (2018). More than 50% or Less than 70% Chance: Pragmatic Implications of Single-Bound Probability Estimates. *Journal of Behavioral Decision Making, 31*(1), 138-150. https://doi.org/10.1002/bdm.2052

Holleman, B. (2006). The meanings of "yes" and "no". An explanation of the forbid/allow asymmetry. *Quality and Quantity, 40*, 1–38. https://doi.org/10.1007/S11135-005-4479-6

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*(4), 341–350. https://doi.org/10.1037/0003-066X.39.4.341

Kelly, M. P., & Barker, M. (2016). Why is changing health-related behaviour so difficult? *Public Health, 136*, 109-116. https://doi.org/10.1016/j.puhe.2016.03.030

Keren, G. (2011). On the definition and possible underpinnings of framing effects: A brief review and a critical evaluation. In G. Keren (Ed.), *Perspectives on Framing*. (pp. 3-33). New York, NY, US: Psychology Press.

Kirchler, E., Maciejovsky, B., & Weber, M. (2005). Framing Effects, Selective Information, and Market Behavior: An Experimental Analysis. *Journal of Behavioral Finance, 6*(2), 90-100. https://doi.org/10.1207/s15427579jpfm0602_4

Kreiner, H., & Gamliel, E. (2018). The Role of Attention in Attribute Framing. *Journal of Behavioral Decision Making, 31*(3), 392-401. https://doi.org/10.1002/bdm.2067

Kühberger, A. (1998). The Influence of Framing on Risky Decisions: A Meta-analysis. *Organizational Behavior and Human Decision Processes, 75*(1), 23-55. https://doi.org/10.1006/obhd.1998.2781

Kühberger, A. (2017). Framing. In R. F. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory, 2nd ed.* (pp. 79-98). New York, NY, US: Routledge.

LeBoeuf, R. A., & Shafir, E. (2003). Deep thoughts and shallow frames; on the susceptibility to framing effects. *Journal of Behavioral Decision Making, 16*(2), 77–92. https://doi.org/10.1002/bdm.433

Levin, I. P. (1987). Associative effects of information framing. *Bulletin of the Psychonomic Society, 25*(2), 85-86. https://doi.org/10.3758/BF03330291

Levin, I. P., & Gaeth, G. J. (1988). How Consumers are Affected by the Framing of Attribute Information Before and After Consuming the Product. *Journal of Consumer Research, 15*(3), 374-378. https://doi.org/10.1086/209174

Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All Frames Are Not Created Equal: A Typology and Critical Analysis of Framing Effects. *Organizational Behavior and Human Decision Processes, 76*(2), 149-188. https://doi.org/10.1006/obhd.1998.2804

Martiny-Huenger, T., Bieleke, M., Doerflinger, J., Stephensen, M. B., & Gollwitzer, P. M. (2020). Deliberation decreases the likelihood of expressing dominant responses. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-020-01795-8

Maule, J., & Villejoubert, G. (2007). What lies beneath: Reframing framing effects. *Thinking & Reasoning, 13*(1), 25-44. https://doi.org/10.1080/13546780600872585

McClung, D. & Schaerer, P.A. (2006). *The Avalanche Handbook*. Seattle, WA: The Mountaineers Books.

McKenzie, C. R. M. (2004). Framing effects in inference tasks—and why they are normatively defensible. *Memory & Cognition, 32*(6), 874-885. https://doi.org/10.3758/BF03196866

McKenzie, C. R. M., & Nelson, J. D. (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin & Review, 10*(3), 596-602. https://doi.org/10.3758/bf03196520

Nelson, T. E., Oxley, Z. M., & Clawson, R. A. (1997). Toward a Psychology of Framing Effects. *Political Behavior, 19*(3), 221-246.https://doi.org/10.1023/A:1024834831093

Payne, J.W., Sagara, N., Shu, S.B. *et al.* (2013) Life expectancy as a constructed belief: Evidence of a live-to or die-by framing effect. *Journal of Risk and Uncertainty*, 46**,** 27–50. https://doi.org/10.1007/s11166-012-9158-0

Peters, E., Hart, P. S., & Fraenkel, L. (2011). Informing patients: the influence of numeracy, framing, and format of side effect information on risk perceptions. *Medical Decision Making, 31*(3), 432-436. https://doi.org/10.1177/0272989x10391672

Piñon, A., & Gambara, H. (2005). A meta-analytic review of framing effect: Risky, attribute and goal framing. *Psicothema, 17*(2), 325-331.

R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rothman, A. J., & Salovey, P. (1997). Shaping perceptions to motivate healthy behavior: The role of message framing. *Psychological Bulletin, 121*(1), 3-19. https://doi.org/10.1037/0033-2909.121.1.3

Scheufele, D.A. and Tewksbury, D. (2007), Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models. *Journal of Communication, 57*(1), 9-20. https://doi.org/10.1111/j.0021-9916.2007.00326.x

Sheeran, P., and Webb, T. L. (2016) The Intention–Behavior Gap. *Social and Personality Psychology Compass*, 10, 503– 518. https://doi.org/10.1111/spc3.12265

Sher, S., & McKenzie, C. R. M. (2006). Information leakage from logically equivalent frames. *Cognition, 101*(3), 467-494. https://doi.org/10.1016/j.cognition.2005.11.001

Sieck, W., & Yates, J. F. (1997). Exposition effects on decision making: Choice and confidence in choice. *Organizational Behavior and Human Decision Processes, 70*(3), 207–219. https://doi.org/10.1006/obhd.1997.2706

Simis, M. J., Madden, H., Cacciatore, M. A., & Yeo, S. K. (2016). The lure of rationality: Why does the deficit model persist in science communication? *Public Understanding of Science, 25*(4), 400-414. https://doi.org/10.1177/0963662516629749

Simon, A.F., Fagley, N.S. and Halleran, J.G. (2004), Decision framing: moderating effects of individual differences and cognitive processing. *Journal of Behavioral Decision Making, 17*, 77-93. https://doi.org/10.1002/bdm.463

Teigen, K. H. (2015). Framing of Numerical Quantities. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell Handbook of Judgment and Decision Making* (pp. 568-589). West Sussex, UK: John Wiley & Sons.

Teigen, K. H., Halberg, A.-M., & Fostervold, K. I. (2007). Single-limit interval estimates as reference points. *Applied Cognitive Psychology, 21*(3), 383-406. https://doi.org/10.1002/acp.1283

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453-458. https://doi.org/10.1126/science.7455683

Weber, M., Keppe, H.-J., & Meyer-Delius, G. (2000). The impact of endowment framing on market prices — an experimental analysis. *Journal of Economic Behavior & Organization, 41*(2), 159-176. https://doi.org/10.1016/S0167-2681(99)00092-X

Wulff, D., Markant, D., Pleskac, T. J., & Hertwig, R. (2019). Adaptive exploration: What you see is up to you. In R. Hertwig, T. Pleskac, T. Pachur, & the Center for Adaptive Rationality (Eds.), *Taming Uncertainty* (pp. 131–152). Boston, MA: MIT Press.

Paper 2

JOURNAL OF RISK RESEARCH
https://doi.org/10.1080/13669877.2021.1905693

Liking and perceived safety across judgments of distinct instances of a category of activity

Matthew B. Stephensen and Torsten Martiny-Huenger

Psychology Department, UiT The Arctic University of Norway, Tromsø, Norway

ABSTRACT

Prior research on the affect heuristic demonstrated that the more a person likes an object or activity, the safer and more valuable it is judged to be. That relation was found when judging stimuli at the categorical level (e.g., nuclear power, airplane travel, heart surgery). Yet risk judgments and decisions usually pertain to specific instances of an object or activity rather than their categorical representations. We examined whether the relation between liking and perceived safety holds across multiple judgments of specific instances of an activity distinguished by contextual information. In four studies ($N = 372$), participants with domain-specific experience (backcountry skiers) completed multicue risk judgments under high uncertainty (judging the avalanche risk in backcountry skiing scenarios) and reported their degree of liking the scenarios. We demonstrate that the positive relation between liking and perceived safety holds across multiple judgments of specific instances of the activity. Furthermore, the liking-perceived safety relation (i.e., judging liked slopes to be safe, judging disliked slopes to be unsafe) held among backcountry skiers who like the activity and consider it safe at the categorical level. We discuss these findings from the perspective that contextual valence and perceived risk can dynamically diverge from categorical valence and perceived risk when perceiving specific instances of that category. These findings have implications for research on attitudes toward risk in extreme sports and other high-risk activities. Although it has been proposed that participants in extreme sports like risk and the thrill it provides, we found that backcountry skiers exhibit a healthy positive relation between liking and perceived safety when judging specific instances of skiing in avalanche terrain.

ARTICLE HISTORY
Received 15 September 2020
Accepted 22 February 2021

KEYWORDS
Affect heuristic; risk judgments; decision making; uncertainty

The affect heuristic across judgments of distinct instances of a category of activity

Standing atop a snow-covered mountain beyond the groomed slopes of a ski resort and the watchful eye of its ski patrol, a backcountry skier contemplating a ski descent is faced with the complex task of judging the risk of avalanche in a highly uncertain environment. Affective processes have been shown to play a role in a range of judgments (Blanchette and Richards 2010; Damasio 1994; Lerner et al. 2015; Loewenstein et al. 2001; Schwarz 2012; Slovic et al. 2004; Wardman 2006; Zajonc 1980) and risk judgments are no exception. Research found that the

CONTACT Matthew Stephensen matthew.stephensen@uit.no; matthew.stephensen@gmail.com Faculty of Health Sciences, Department of Psychology, UiT The Arctic University of Norway, Postboks 6050 Langnes, Tromsø, 9037, Norway.

more a person likes an object or activity, the safer and more valuable it appears, whereas the more a person dislikes an object or activity, the more unsafe and less valuable it appears (Alhakami and Slovic 1994; Finucane et al. 2000; Slovic et al. 2002). The positive relation between liking and perceived safety was found when judging categorical representations of objects (e.g., chemical plants, pesticides, nuclear power, and menopause drugs) and activities (e.g., cigarette smoking, heart surgery, airplane travel, and surfing) independent of specific contextual information, details, or constraints. Although our skier likes to ski in the backcountry and generally considers it a safe activity at the categorical level, sentiments that have certainly influenced her decision to be atop the mountain, she must now selectively attend to affective and cognitive information cues to judge the risk of her specific situation. Does the positive relation between liking and perceived safety (alternatively, disliking and perceived danger) found when judging categorical representations hold when our skier perceives a specific instance of backcountry skiing distinguished by contextual details? What the research on the liking-perceived safety relation has not adequately examined is whether that relation is dependent upon the level at which the target of judgment is perceived, be it the abstract, general level of categorical representations (e.g., backcountry skiing) or at the level of a distinct, context-specific instance of that category (e.g., skiing a specific mountain under certain conditions).

Both liking and safety can be judged either at the level of a distinct instance of an activity or at the level of a prototypical, categorical representation of that activity (Medin 1989; Yee and Thompson-Schill 2016). For example, the *category* valence of an activity is the positive or negative affective response evoked when considering a general, categorical representation of that activity (e.g., I like backcountry skiing). By contrast, we define *contextual* valence as the affective response evoked when perceiving a specific instance of that activity that is distinguished by context-specific information (e.g., I like to ski this specific mountain under these particular conditions). Research on approach-avoidance motivations and emotions found that the categorical valence of a stimulus that is (relatively) isolated from a contextual situation and the contextual motivational valence of a stimulus that is determined by contextual affordances and constraints are distinct evaluative responses that do not necessarily converge (Elliot, Eder, and Harmon-Jones 2013; Moors and De Houwer 2001). This evidence raises an important issue concerning liking, perceived safety, and the relation between the two. Although it is reasonable to assume that affective evaluations and risk perceptions can similarly differ between levels of perception, what is presently unclear is if both contextual valence and contextual perceived risk diverge from categorical perspectives in a way that defies or upholds the positive relation between liking and perceived safety.

Category-level and context-level judgments can diverge in various ways that have implications for the robustness of the relation between liking and perceived safety. It is possible that judgments of specific instances of an object or activity defy the positive relation between liking and perceived safety found for judgments of categorical representations. For example, our backcountry skier who likes that category of activity and generally considers it to be safe could maintain that category-level liking and have a similarly positive affective evaluation of a specific instance of backcountry skiing despite judging it unsafe. Or she could dislike a specific instance of backcountry skiing while maintaining her category-level perspective of safety, thereby judging the disliked instance to be safe. Both cases represent context-level judgments diverging from categorical perspectives in a way that defies the positive relation between liking and perceived safety. Alternatively, the liking-perceived safety relation found for judgments at the categorical level could hold across multiple distinct instances, despite both contextual judgments failing to converge with the established valence and perceived safety of that activity at the categorical level. Our skier might dislike a specific instance of an activity and perceive it to be unsafe despite generally liking that category of activity and considering it safe. In the present research, we tested evidence for the latter reasoning that would indicate that the positive relation between liking and perceived safety holds across judgments of specific, distinct instances of an activity.

The conceptualization of affect

The affect heuristic theory is one of several theories that explain the role of affective evaluations in guiding judgments and decisions. The affect heuristic theory proposes that feelings of goodness or badness reflect the positive and negative qualities associated with a stimulus. Such affective responses occur rapidly and automatically before more elaborate cognitive processes occur. Consequently, people rely on these affective evaluations as a valuable and compelling orienting mechanism for quick and efficient judgments (Finucane et al. 2000; Slovic et al. 2002, 2004). Other theoretical frameworks that describe the role of affect in judgments and decisions are, for example, the feelings-as-information hypothesis (Schwarz 2012; Schwarz & Clore, 1983, 2007), the risk-as-feelings hypothesis (Loewenstein et al. 2001), and the somatic marker hypothesis (Damasio 1994). A key point on which these theories all agree is that the affective response to a stimulus influences the ensuing judgment of that stimulus. In our present research, the affect heuristic theory serves as a representative case of those various theories describing the role of affect in guiding judgments and decisions. We privilege the affect heuristic theory because our conceptualization of affect reflects the way in which it is regarded in the affect heuristic theory. The various theories describing the influence of affect on judgments and decisions disagree about how affect is conceptualized, with several of the theories regarding affect as equivalent to emotions or feelings (for a detailed discussion, see Wardman 2006). By contrast, Slovic and colleagues (2004) regard affect as "*a faint whisper of emotion*" (p. 312) rather than a complex range of visceral emotions. In accordance with the affect heuristic theory, we conceptualize affect as a generalized positive or negative response to the target of judgment. Although we align our research with the affect heuristic's conceptualization of affect, our findings are relevant for the various existing theories on the role of affect in guiding risk judgments and decisions.

The present studies

Backcountry skiing in avalanche terrain is a useful context for examining if the relation between liking and perceived safety holds when judging specific instances of an activity. We interpret backcountry skiers' prior behavior of going backcountry skiing as indication that they like the activity and consider it safe at the categorical level. This reflects the positive relation between liking and perceived safety that Alhakami and Slovic (1994), Finucane et al. (2000), and Slovic et al. (2002) proposed under the affect heuristic theory. What is important for our investigation, however, is that skiing in avalanche terrain is objectively risky under certain circumstances (Engeset et al. 2018; Niedermeier et al. 2019). Avalanche terrain is a complex and highly uncertain environment, where poor decisions are seldom marked by valid feedback, making it extremely difficult for even the most experienced backcountry skier to judge avalanche risk (Ebert 2019; Hogarth, Lejarraga, and Soyer 2015; Zweifel and Haegeli 2014). It is therefore possible to examine whether objectively uncertain or dangerous instances of backcountry skiing are disliked and perceived as unsafe, in accordance with the liking-perceived safety relation, despite the population liking the activity and perceiving it as safe at the categorical level.

Across four studies, we presented participants with a series of specific instances of backcountry skiing scenarios. We employed a method similar to that used by Alhakami and Slovic (1994) of directly measuring valence and perceived safety. Participants judged the safety of each scenario in terms of the avalanche risk, henceforth referred to as *judged safety*, and reported their degree of liking each scenario, henceforth referred to as *scenario liking*. We define liking as a generalized positive or negative affective response associated with a stimulus (Alhakami and Slovic 1994; Finucane et al. 2000; Winkielman, Zajonc, and Schwarz 1997; Zajonc 1980). Through our use of the term liking, we distinguish our valence-based measurement of the positive-negative evaluative dimension of affect from more emotional responses such as happiness and

sadness (see Sjöberg 2006, and Wardman 2006 for a more detailed discussion). Participants in Study 1 included a mix of student participants and backcountry skiers in order to obtain a broad range of participant experience. The participant samples for Studies 2, 3, and 4 were exclusively recruited from among active backcountry skiers.

In addition to examining the liking-perceived safety relation when judging specific instances, we attempted to test the causal direction associated with the affect heuristic (see Finucane et al. 2000; Slovic et al. 2002; Slovic et al. 2004). To test whether liking influenced judged safety, we included a within-subject manipulation in each of the studies with the aim of manipulating scenario liking while holding constant the objective risk level of each scenario. We attempted to manipulate the attractiveness of the scenarios by subtly including cues that the slopes were either untracked or had recently been skied (Studies 1-3), or through a more salient manipulation of presenting the slopes in sunny, high contrast or cloudy, low contrast conditions (Study 4). Anecdotal evidence (McCammon 2002, 2004) suggests that backcountry skiers value access to untracked slopes and should therefore prefer scenarios without tracks. As for the weather manipulation, the results of pre-testing indicated that scenario photos with a background of sunny, clear blue skies were preferred to otherwise identical scenario photos with a background of overcast, cloudy grey skies. Unfortunately, the manipulations did not successfully influence scenario liking and we were unable to test for a causal effect. Nonetheless, this failure provides valuable insight into the theoretical framework that dominates avalanche research and education, to which we return in the General Discussion. Importantly, the failed manipulations did not adversely affect our main focus of presenting evidence of the positive relation between liking and perceived safety.

We confirm that at the time of writing, the four studies reported here are all the studies we conducted on the relation between liking and perceived safety. We report all measurements assessed and all manipulations implemented in each study. These studies were conducted in accordance with the ethical research protocols of UiT The Arctic University of Norway and the Norwegian Center for Research Data (NSD). Study 4 was pre-registered. The data, R script for data processing and analysis, the pre-registration of Study 4, and the scenarios used in the studies are available on the Open Science Framework (https://doi.org/10.17605/OSF.IO/VA28N).

Study 1

In an internet-based study, a mixed sample of participants were presented four scenarios of backcountry skiing in avalanche terrain. Participants reported judged safety and scenario liking for each scenario.

Method

Participants. Forty-six participants (18 self-identified as male, 28 as female, $M_{age} = 26$, range 18-52, $SD = 8.34$) were recruited from a psychology course (63% of participants), receiving course credit for participation, and recruited from among backcountry skiers in Norway (37% of participants). We did not conduct a priori power analysis because of the difficulty of estimating power for linear mixed models (Johnson et al. 2015; Westfall, Kenny, and Judd 2014). We instead set the minimum sample size at 40 and recruited as many participants as possible within a predefined 4-week period for data collection. We did not commence analysis before completing data collection. Participants could complete the study in Norwegian or English.

Materials. We conducted the study online and used the jsPsych programme (de Leeuw 2015) to control the stimulus presentation in the web browser. We developed six hypothetical scenarios each depicting a distinct backcountry ski descent on a snow-covered mountain slope (i.e., avalanche terrain, see Figure 1 for an example). Each scenario began with a description of the

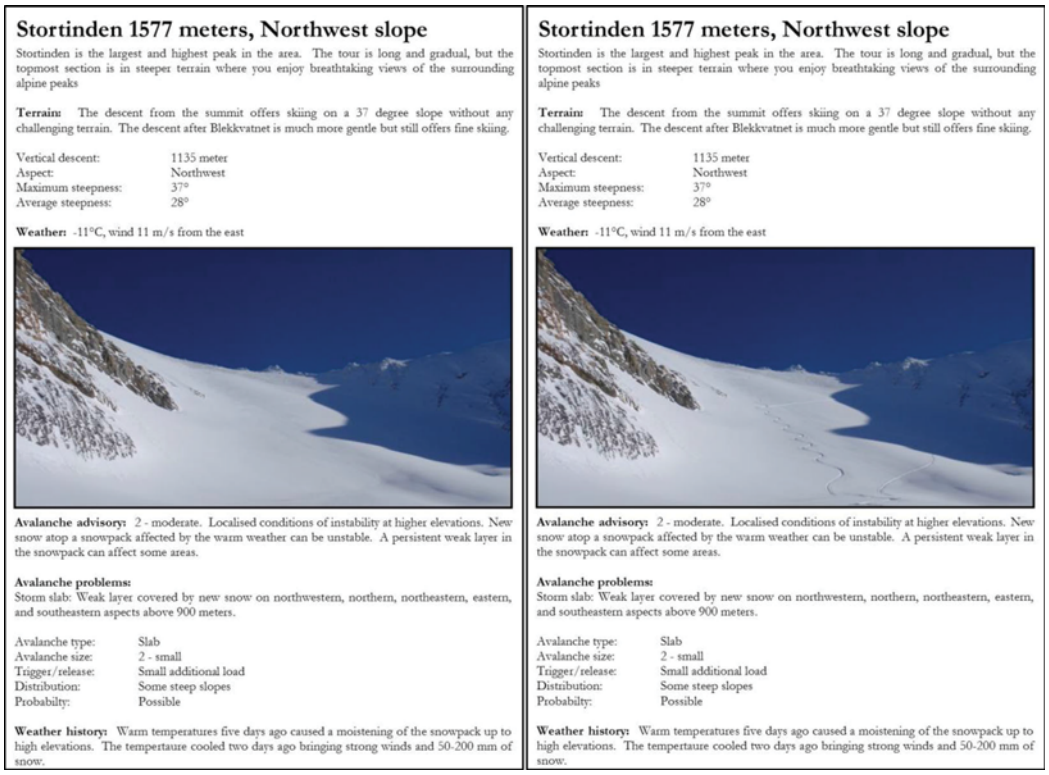


Figure 1. Example of the backcountry skiing scenarios judged by participants. There were two versions of each scenario: one version included ski tracks in the photo whereas the photo in the second version was untracked. Participants were presented with only one version of the scenario.

tour that included basic information on the terrain, elevation, average and maximum slope steepness, and the current weather; information that is relevant for judging the avalanche risk. We presented the tour description according to the format and content of local backcountry skiing guidebooks. A photograph of the mountain slope was located below the tour information. Beneath the photograph was the forecasted avalanche danger, the current avalanche problem(s) caused by the combination of weather and snow conditions, and the weather history. We based the content and format of the avalanche hazards information on historic avalanche forecasts from the Norwegian Avalanche Warning Service (Varsom, n.d.). We did not include any extreme indicators of either very low or very high risk levels in the scenarios; by design, all scenarios were highly uncertain. We prepared two versions of each of the scenarios—one version included the clearly visible ski tracks on the mountain slope in the scenario photograph, whereas there were no ski tracks in the photograph of the other version (see Figure 1 for an example).

Procedures. Participants accessed the online study via a web browser at a time and device of their choosing. All participants indicated their informed consent to participate by first checking a confirmation box and then clicking a button to proceed to the study. They were then instructed to read the scenario information and answer the questions for each scenario. The questions were presented immediately below each scenario and participants viewed them by scrolling down the web page. All the questions for each scenario were presented simultaneously on the same page and could be answered in any order.

For the safety judgments, participants answered three questions following the general statement "*Regarding the avalanche risk…*": 1) "*Is it safe to ski the slope in these conditions?*" 2) "*Is it dangerous to ski the slope in these conditions?*" 3) "*Is the snowpack stable enough to ski this slope?*" on a 7-point scale labeled "*Not at all*" and "*Very much*" at the extreme points. For the degree of liking the

scenario, following the general statement "*Regarding your desire to ski this slope, the slope appears…*" participants rated the scenario according to the following three adjectives: 1) *"Attractive", 2)* "*Uninteresting"*, and *3)* "*Enjoyable"* on a 7-point scale labeled "*Not at all*" and "*Very much*" at the extreme points. A third set of three questions concerned participants' degree of confidence in their safety judgment. We measured confidence for purposes unrelated to the focus of the current article and do not investigate or discuss the measure here. Participants were required to answer all questions on judged safety, scenario liking, and confidence to proceed with the study.

Each participant judged four scenarios. The selection of four scenarios from the available six, which scenarios were presented with or without tracks, and the order of presentation were randomly determined for each participant. Upon completing four scenarios, participants reported their age and gender. Participants then reported their skiing ability, avalanche training, years of backcountry skiing experience and average number of backcountry skiing days per season, each of which was measured on a 7-point scale. For exploratory purposes unrelated to the current article, we measured participants' past exposure to avalanches and their use of backcountry ski guides. Upon completing these questions, participants were asked to answer an open question about the purpose of the study to check if they had identified the manipulation; no participant identified the manipulation.

Data preparation and analysis. We calculated mean scores for judged safety (3 items, the question on danger reverse-coded, Cronbach's alpha = .90, $N = 184$, $M = 3.45$, $SD = 1.40$) and for scenario liking (3 items, the question on uninteresting reverse-coded, Cronbach's alpha = .92, $N = 184$, $M = 4.59$, $SD = 1.63$) per participant and scenario. We calculated a mean experience score per participant from the measurements of skiing ability, avalanche training, years of backcountry skiing experience and average number of backcountry skiing days per season (4 items, Cronbach's alpha = .92, N = 46, $M = 3.21$, $SD = 1.88$). Two participants did not respond to all four measures of experience. We calculated their mean experience scores using the available measures for those two participants (Schafer and Graham 2002).

We used R (R Core Team 2017) and the *lme4* package (Bates et al. 2015) to fit linear mixed models to predict judged safety, estimated using maximum likelihood and Nelder-Mead optimization. We included intercepts for participants and scenarios as varying effects, thereby accounting for by-subject and by-scenario variability. We report the intraclass correlation (ICC) for the varying effect participant as an indication of the amount of variance in judged safety accounted for by individual difference between participants. Similarly, we report the ICC for the varying effect scenario as an indication of the amount of variance in judged safety accounted for by objective differences between scenarios. We used the *lmerTest* package (Kuznetsova, Brockhoff, and Christensen 2017) with Satterthwaite approximations to obtain *p*-values. After fitting the regression models, we analyzed the observations ($N = 184$, 1 observation per scenario per participant) for outliers using the *LMERConvenienceFunctions* package (Tremblay and Ransjin 2020). We excluded three outlier observations with a standardized residual value greater than 2.5 standard deviations from 0.

Results and discussion

Scenario liking predicted judged safety, $b = .46$, $SE_b = .06$, *95*% *CI* [.34, .58], $p < .001$, with an effect size *std.* $b = .53$. Consistent with the findings from prior research (Alhakami and Slovic 1994; Finucane et al. 2000; Slovic et al. 2002), scenario liking positively relates to judged safety. The more a participant liked a backcountry ski tour, the higher that participant judged avalanche-related safety. Importantly, in contrast to prior research in which participants judged categorical representations of objects and activities, we found this relation across multiple judgments of specific instances of an activity for a range of scenario liking (1.00 to 7.00) and judged safety (1.00 to 6.33) scores. 27.9% of variation in judged safety was attributable to the

difference between participants, χ^2 (1) = 15.55, indicating that participants' safety judgments were not stable but differed (72.1% within participant variance) between judgments. Moreover, 2.9% of variation in judged safety was attributable to the difference between scenarios, χ^2 (1) = 2.35, indicating that each scenario did not elicit a stable judgment of safety. Judged safety varied greatly for each scenario according to subjective interpretation of the characteristics.

Student participants had less experience with both the activity depicted in the scenarios and the judgement task (63% of the sample, student subgroup $M_{experience}$ = 2.13 with a possible range from 1 to 7) than the backcountry skier participants did (37% of the sample, skier subgroup $M_{experience}$ = 5.01). Yet analysis for a moderation effect by experience indicated that an interaction between scenario liking and experience, $b = -.02$, $SE_b = .03$, *95% CI* [-.09, .04], $p = .518$, did not predict judged safety. There was no evidence that the relation between scenario liking and judged safety differed between the less experienced student participants and the more experienced backcountry skiers. Finally, the presence or absence of tracks did not influence scenario liking, $b = .00$, $SE_b = .16$, *95% CI* [-.31, .33], $p = .965$, or judged safety, $b = -.14$, $SE_b = .15$, *95% CI* [-.44, .15], $p = .344$. The failure of this manipulation to affect scenario liking made it impossible to test for causality in the liking-perceived safety relation. We next sought to replicate these findings in a study with a sample of exclusively backcountry skiers for whom the categorical valence and categorical perceived safety of the activity are established.

Study 2

We tested the relation between scenario liking and judged safety with a non-student sample of exclusively backcountry skiers who like the activity and judged it to be safe and valuable at the categorical level. We conducted this study with the identical design, materials and procedures used in Study 1. In what follows, we only report the unique aspects of this replication.

Participants. Fifty-four participants (41 self-identified as male, 13 as female, M_{age} = 32, range 17-54, SD = 9.68) were recruited among backcountry skiers in Norway. We announced the study to the attendees at two avalanche safety seminars in February 2018, inviting them to participant at any time during the following 3 weeks. The minimum sample size was set at 40, as per Study 1. There was no upper limit on the number of participants in the study; we recruited as many participants as possible during the predefined 3-week period for data collection. Participants had on average more experience with backcountry skiing and avalanche safety judgments than participants in Study 1 (Study 2, $M_{Experience}$ = 4.96 with a possible range from 1 to 7, as compared to Study 1, $M_{Experience}$ = 3.21). Notably, the topic of the seminars from which we recruited participants was improved decision-making in avalanche terrain, where they were instructed on the correct methods and potential errors when judging avalanche risk. We did not commence analysis before completing data collection. The study was conducted in Norwegian.

Data preparation and analyses. We prepared and analyzed the data according to the same methods reported for Study 1. We calculated mean scores for judged safety (Cronbach's alpha = .92, N = 216, M = 3.34, SD = 1.43) and for scenario liking (Cronbach's alpha = .89, N = 216, M = 5.29, SD = 1.50) per participant and scenario, and a mean experience score (Cronbach's alpha = .77, N = 54, M = 4.96, SD = 1.11) per participant. Two observations were identified as outliers (standardized residual value greater than 2.5 standard deviations from 0) and removed from the data.

Results

Consistent with the results of Study 1, scenario liking predicted judged safety, $b = .18$, $SE_b = .07$, *95% CI* [.05, .30], $p = .008$, with an effect size *std.* $b = .20$. Scenario liking positively corresponds to judged safety across multiple judgments of specific instances of an activity, for a range of

scenario liking (1.00 to 7.00) and judged safety (1.00 to 7.00) scores. 15.2% of variation in judged safety was attributable to the difference between participants, χ^2 (1) = 6.34, while 5.5% of variation in judged safety was attributable to the difference between scenarios, χ^2 (1) = 5.71. Judged safety was not stable for each participant. Nor was it stable for each scenario. Analysis to test for a moderation effect by experience indicated that an interaction between scenario liking and experience, $b = -.09$, $SE_b = .06$, *95% CI* [-.22, .03], $p = .144$, did not predict judged safety. Overall, the results of Study 2 provide further evidence of a positive relation between scenario liking and safety judgments at the contextual level, qualitatively replicating the results of Study 1. Finally, unlike in Study 1, the presence of tracks increased scenario liking, $b = .30$, $SE_b = .15$, *95% CI* [.00, .60], $p = .050$, contrary to the direction of effect we predicted for the manipulation. Despite that effect on scenario liking, neither the main effect tracks, $b = .09$, $SE_b = .17$, *95% CI* [-.24, .43], $p = .591$, nor an interaction between scenario liking and the tracks manipulation, $b = .07$, $SE_b = .12$, *95% CI* [-.17, .30], $p = .569$, predicted judged safety. To investigate whether the smaller effect size in Study 2 was in any way specific to the population of backcountry skiers, we conducted a third study with a sample of exclusively backcountry skiers in a different country using new scenarios adjusted for that new population.

Study 3

We conducted a direct replication of the previous study with a sample of exclusively backcountry skiers recruited from a different population. Participants in Studies 1 and 2 were recruited in Norway, reflecting a Scandinavian perspective on backcountry skiing and avalanche risk. Participants in Study 3 were recruited from the USA, reflecting a North American perspective on backcountry skiing and avalanche risk. Otherwise, we conducted this study with the identical design and procedures used in Studies 1 and 2 using new scenarios that were conceptually the same but adjusted to the norms of the population. In what follows, we only report the unique aspects of this replication.

Participants. Forty-one participants (29 self-identified as male, 12 as female, $M_{age} = 26$, range 15-50, $SD = 7.85$) were recruited via an email announcement sent to backcountry skiers in the western USA in March 2018. We obtained the email addresses from a registry of individuals who, when registering to attend an avalanche seminar, indicated their willingness to participate in studies on avalanche safety. Participants who completed the study were eligible to register for a prize draw to win one of six USD 50 gift certificates for an online store. The minimum sample size was set at 40, as per Studies 1 and 2. We recruited as many participants as possible during a predefined 4-week period for data collection. We completed all data collection before beginning analysis. The study was conducted in English.

Materials. We used six new scenarios of the same design as those previously reported, but with new content suited to the norms of the target population. We developed six hypothetical scenarios using measurement units (e.g., Fahrenheit), geography, tour descriptions and photographs familiar to a population in the western USA. We changed the scenario photos to depict terrain similar to that of the Rocky Mountain region from where participants were recruited. We based the tour descriptions – both content and language – on descriptions found in American backcountry skiing guidebooks for that region. We based the weather, avalanche danger forecast and avalanche problems in each scenario on historic avalanche forecasts from local avalanche warning services (Colorado Avalanche Information Center, n.d.; Gallatin National Forest Avalanche Center, n.d.; Utah Avalanche Center, n.d.). A senior avalanche researcher at Montana State University reviewed all the scenarios to ensure that their content was suitable for the target population. We did not include any extreme indicators of either very low or very high risk levels in the scenarios. By design, the avalanche risk in all scenarios was uncertain.

Data preparation and analyses. We prepared and analyzed the data according to the same procedures reported for Studies 1 and 2. We calculated mean scores for judged safety (alpha = .93, $N = 164$, $M = 3.07$, $SD = 1.43$) and for scenario liking (Cronbach's alpha = .84, $N = 164$, $M = 5.46$, $SD = 1.29$) per participant and scenario, and a mean experience score (Cronbach's alpha = .60, $N = 41$, $M = 3.92$, $SD = 1.00$) per participant. One participant did not answer all questions measuring experience and we calculated the mean experience score for that participant using the available measures. Two observations were identified as outliers (standardized residual value greater than 2.5 standard deviations from 0) and removed from the data.

Results

Consistent with the results of Studies 1 and 2, scenario liking predicted judged safety, $b = .15$, $SE_b = .08$, *95% CI* [.00, .30], $p = .050$, with an effect size *std.* $b = .18$. Scenario liking positively relates to judged safety across multiple judgments of specific instances of an activity, for a range of scenario liking (1.00 to 7.00) and judged safety (1.00 to 6.33) scores. 20.5% of variation in judged safety was attributable to the difference between participants, χ^2 (1) = 13.82, while 27.0% of variation in judged safety was attributable to the difference between scenarios, χ^2 (1) = 51.74. Judged safety was not stable for each participant. Nor was it stable for each scenario. Analysis to test for a moderation effect by experience indicated that an interaction between scenario liking and experience, $b = .04$, $SE_b = .07$, *95% CI* [-.09, .19], $p = .502$, did not predict judged safety. The results of Study 3 provide further evidence of a positive relation between scenario liking and safety judgments at the contextual level, replicating the results of Studies 1 and 2. Replicating this result within a different population and with new materials indicates that the results of the previous two studies were not unique to the population or to the materials and manner in which they were presented in Studies 1 and 2. However, although we adjusted the scenarios used in Study 3 so that their content would be suited to the target population, it is possible that any unforeseen mismatch between scenarios and the real-world decision environments that are familiar to the participants could limit the comparability and generalizability of the studies.

The presence of tracks did not influence scenario liking, $b = .12$, $SE_b = .14$, *95% CI* [-.16, .40], $p = .394$. However, unlike previous studies, the presence of tracks increased judged safety, $b = .31$, $SE_b = .15$, *95% CI* [.00, .61], $p = .049$. Moreover, there is some evidence that an interaction effect between scenario liking and tracks predicted judged safety: scenario liking, $b = .04$, $SE_b = .09$, *95% CI* [-.14, .23], $p = .645$; tracks, $b = -.96$, $SE_b = .69$, *95% CI* [-2.32, .40], $p = .170$; and their interaction, $b = .23$, $SE_b = .12$, *95% CI* [-.01, .48], $p = .064$. This suggests that the relation between scenario liking and judged safety was stronger for scenarios with tracks. Nonetheless, we were unable to examine causality because our manipulation did not affect scenario liking as expected.

Study 4

In Studies 1, 2, and 3, the questions on judged safety were presented before the questions on scenario liking. To test for an order effect, in Study 4 we counterbalanced the order of the liking and safety judgments so that half of the participants judged safety first while the other half judged liking first. Moreover, after the inconsistent effect of the tracks manipulation on scenario liking and judged safety in Studies 1 to 3, we used a more salient manipulation of weather in an attempt to affect scenario liking to test for a causal effect on judged safety. Pre-testing indicated that scenario photos with a background of sunny, clear blue skies were preferred to otherwise identical scenario photos with a background of overcast, cloudy grey skies.

Methods

Participants. Two-hundred and thirty-one participants (162 self-identified as male, 68 as female, 1 as other; $M_{age} = 35$, range 19-62, $SD = 10.06$) were recruited via email announcements sent to backcountry skiers in Norway in January 2020. We obtained the email addresses from a registry of individuals who, when registering for an avalanche seminar, indicated their willingness to participate in studies on avalanche safety. Participants who completed the study were eligible to register for a prize draw to win one avalanche airbag and air cylinder. The minimum sample size was set at 105 based on a priori simulation-based power analysis using the smallest effect size measured in Studies 1, 2, and 3. We recruited as many participants as possible during a predefined 4-week period for data collection. We completed all data collection before beginning analysis. Participants could complete the study in English or Norwegian.

Materials and procedures. We developed four scenarios of the same design as those previously reported (all scenario photos were without tracks) for Studies 1 to 3. We prepared two versions of each scenario: the photo in one version had a sunny, clear blue sky and the mountain was brighter and in higher contrast, whereas the photo in the other version had an overcast, cloudy grey sky and the mountain was darker and in lower contrast. Each participant in the study judged four scenarios. The only fixed aspect was the ratio of two sunny scenarios and two cloudy scenarios per participant. Whether the scenarios were sunny or cloudy and their order of presentation were randomly determined for each participant. As per Studies 1 to 3, the three safety judgment questions and the three liking judgment questions were presented together on the same page below the scenario. However, the order of those question blocks was counterbalanced between participants. The three questions on rated liking were presented above the three questions on rated safety for even-numbered participants, and presented below the three questions on rated safety for odd-numbered participants.

Data preparation and analyses. We prepared and analyzed the data according to the same procedures reported for Studies 1 to 3. We calculated mean scores for judged safety (Cronbach's alpha = .88, $N = 924$, $M = 3.07$, $SD = 1.29$) and for scenario liking (Cronbach's alpha = .88, $N = 924$, $M = 4.94$, $SD = 1.54$) per participant and scenario, and mean experience score (Cronbach's alpha = .73, $N = 231$, $M = 4.53$, $SD = 1.08$) per participant. Nine observations were identified as outliers (standardized residual value greater than 2.5 standard deviations from 0) and removed from the data.

Results

Consistent with our previous results, scenario liking predicted judged safety, $b = .22$, $SE_b = .03$, *95% CI* [.17, .28], $p < .001$, with an effect size *std. b* = .26. A higher value of scenario liking relates to a higher value of judged safety across multiple judgments of specific instances of an activity, for a range of scenario liking (1.00 to 7.00) and judged safety (1.00 to 6.33) values. This replicates the results of Studies 1 to 3. 29.1% of variation in judged safety was attributable to the difference between participants, χ^2 (1) = 89.59, while 2.1% of variation in judged safety was attributable to the difference between scenarios, χ^2 (1) = 17.42. Judged safety was not stable for each participant. Nor was it stable for each scenario. Analysis for a moderation effect by experience indicated that an interaction between scenario liking and experience, $b = -.03$, $SE_b = .02$, *95% CI* [-.09, .01], $p = .147$, did not predict judged safety. Analysis for an effect from the order of the questions revealed weak evidence that the order of the questions influenced judged safety, $b_{order} = .21$, $SE_b = .11$, *95% CI* [-.01, .43], $p = .067$, effect size *std. b* = .16. Judged safety was on average higher when safety was judged before scenario liking. More importantly, however, an interaction between scenario liking and question order did not predict judged safety, $b = -.02$, $SE_b = .06$, *95% CI* [-.13, .09], $p = .709$. There was no evidence that the order of the questions influenced the magnitude of the relation between scenario liking and judged safety. Finally,

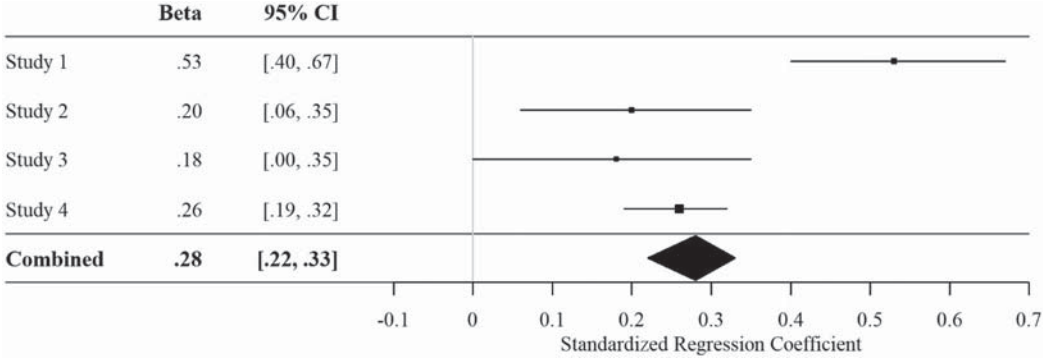


Figure 2. Standardized regression coefficients and 95% CI for scenario liking predicting judged safety for Studies 1-4 and the combined data.

despite being more salient than the subtle tracks manipulation used in Studies 1 to 3, the weather condition of the scenario photo did not influence judged safety, $b = .00$, $SE_b = .07$, *95% CI* [-.14, .14], $p = .988$, or scenario liking, $b = .01$, $SE_b = .07$, *95% CI* [-14., .15], $p = .915$. Failing to manipulate scenario liking as expected, we were again unable to test the causal direct of the relation between affective evaluations and risk judgments.

Synthesis of evidence across studies 1 to 4

We used the meta-analytic *Q* test with studies as a fixed effect to assess the magnitude of variation in the effect sizes across the studies (Schauer and Hedges 2020). Although the effect parameters in all four studies are all in the same direction (i.e., there is a positive relation between liking and judged safety) with *p*-values equal to or less than the conventional inference threshold of .05 for null hypothesis significance testing, the *Q* test revealed evidence of heterogeneity of effect size, χ^2 (3) = 15.42, $p = .002$, across the studies. We measured a larger effect size in Study 1 (see Figure 2). To synthesize the evidence across all studies, we combined individual participant data from the four studies (372 participants, 1472 observations) for pooled analysis to more accurately estimate the effect parameter of the relation between liking and judged safety (da Costa and Sutton 2019). As described for Study 1, we fitted linear mixed models using maximum likelihood to predict the outcome variable judged safety. To account for the heterogeneity of effect parameters between studies, we assigned a varying intercept for studies when estimating all models (in addition to varying intercepts for participants and scenarios). Analysis of the combined data indicates that scenario liking predicted judged safety with an effect size *std.* $b = .28$, $SE_b = .03$, *95% CI* [.22, .32], across all four studies. These effect parameters are identical to those obtained by calculating the average weighted effect size using the effect parameter results from each study. 23.9% of variation in judged safety was attributable to the difference between participants, χ^2 (1) = 128.54, 13.2% of variation in judged safety was attributable to the difference between scenarios, χ^2 (1) = 90.70, and 1.1% of variation in judged safety was attributable to the difference between studies (apart from the different scenarios), χ^2 (1) = 1.00.

One might reasonably assume that the larger, heterogeneous effect size in Study 1 was caused by 63% of participants being students who had less experience with the activity depicted in the scenarios and the risk judgment task. We therefore included experience in the pooled analysis. We fitted a model with the effects scenario liking, participant experience, and their interaction to assess whether it predicted judged safety. Multicollinearity between scenario liking and participant experience was high (VIF = 18.80) so we standardized the predictor and response variables. The combined data indicate that an interaction between scenario liking and participant

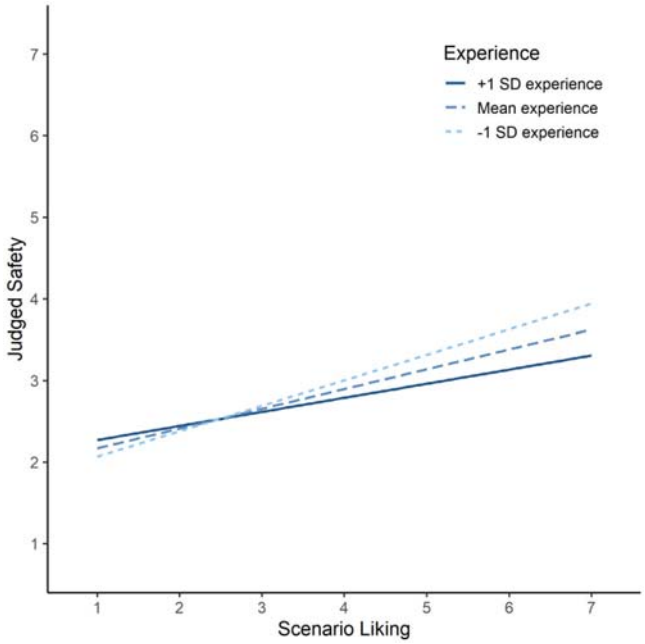


Figure 3. Scenario liking as predictor of judged safety at +1 SD, mean, and -1 SD participant experience scores.

experience predicted judged safety: scenario liking, $b = .28$, $SE_b = .03$, *95% CI* [.23, .33], $p < .001$; experience, $b = -.13$, $SE_b = .03$, *95% CI* [-.20, −.07], $p < .001$; and their interaction, $b = -.08$, $SE_b = .02$, *95% CI* [-.13, −.03], $p = .001$. 23.1% of variation in judged safety was attributable to the difference between participants, χ^2 (1) = 119.94, 13.9% of variation in judged safety was attributable to the difference between scenarios, χ^2 (1) = 91.83, and 0.0% of variation in judged safety was attributable to the difference between studies (apart from the different scenarios), χ^2 (1) = 0.00. There is evidence in the pooled data that participant experience moderates the relation between scenario liking and judged safety: the magnitude of that relation was lower when participant experience was higher (see Figure 3). However, the size of the interaction effect is extremely small.

General discussion

Across four studies, we found evidence that higher self-reported liking of backcountry ski scenarios corresponded to judgments of higher avalanche safety. This aligns with earlier research using judgments of categorical representations of stimuli (Alhakami and Slovic 1994; Finucane et al. 2000; Slovic et al. 2002). Our present research extends those prior findings by demonstrating that the liking-perceived safety relation holds across multiple judgments of highly uncertain, specific instances of an activity that are distinguished by contextual information. Furthermore, the liking-perceived safety relation was found to hold when individuals for whom the activity has established categorical valence and safety judged multiple distinct instances of that activity. Despite the facts that backcountry skiers like backcountry skiing in avalanche terrain and deem the activity to be safe, specific instances of backcountry skiing assumed a negative contextual valence and were perceived as unsafe in accordance with the liking-perceived safety relation. Echoing findings from the field of approach-avoidance motivation and emotion (Elliot, Eder, and Harmon-Jones 2013; Moors and De Houwer 2001), our results show that categorical valence and contextual valence (and, similarly, categorical perceived safety and contextual perceived safety) are distinct judgment processes that do not necessarily converge. The relation between liking

and perceived safety holds at the contextual level of perception, even when the valence and perceived safety of a specific instance of a stimulus conflicts with the established valence and perceived safety of that category of stimulus for the decision maker.

All experiences of phenomena are marked, to varying degrees, with affect. Those positive and negative affective markers are aggregated to create an "affect pool" that provides an affective frame of reference for interpreting any new phenomena a decision maker encounters (Slovic et al. 2004; see also Schwarz 2007, for a similar argument in the context of attitude construction). A categorical representation has an affective value or valence that reflects the aggregated positive and negative markers of prior experiences of specific instances of that category of phenomenon. In that way, a categorical representation is a prototypical expression of the affective frame of reference. By contrast, a distinct, context-specific instance of that category will be marked by a unique array of affective characteristics. The affective response that that distinct instance of a phenomenon evokes depends upon the salience of those affective characteristics and the ease with which the decision maker interprets or maps them according to the affective frame of reference for that category of phenomena (Slovic et al. 2004; Wardman 2006; Wilson and Arvai 2006). A specific instance of an affect-rich phenomenon such as backcountry skiing can include certain affective characteristics that a decision maker does not include in the mental image of a categorical representation, or may lack other affective characteristics typically associated with the categorical representation. When those differences in affective characteristics are great enough, the valence of a distinct, context-specific instance of a phenomenon will diverge from the valence of that category of phenomena. Risk perceptions can differ between levels of perception in similar manner. Because the salient affective qualities and risk characteristics of each scenario differed (to lesser or greater degrees) from participants' categorical representations of backcountry skiing, contextual valence and risk perception diverged from categorical valence. This resulted in dynamic changes between contextual judgments of specific instances.

The focus of our research on contextual judgments of specific instances of a phenomenon is not a methodological contrivance. We believe that such contextual judgments are a common aspect of daily life and, as such, are ecologically valid representations of real-world decisions. Although judging the risk, benefit, and degree of liking an activity in general – such as backcountry skiing – is a valid and realistic judgment task, it is a very different task from judging the risk, benefit, and degree of liking a specific instance of that activity. The salience of affective qualities and risk considerations will be different between the two judgment tasks, despite one target of judgment being a categorical representation of the other. That contextual judgments should diverge from judgment of categorical representations, and differ between distinct contextual judgments, is both natural and beneficial. When faced with a specific potential hazard and high uncertainty, people should ideally decide behavior based on a contextual judgment rather than their general, category-level orientation. Otherwise, decision makers would find themselves trapped by their category-based judgments, doomed to repeat the affective response and risk perception irrespective of contextual characteristics.

The moderating effect of experience

We found evidence that a decision maker's prior experience with the stimulus activity and risk judgment task appears to moderate the relation between affective evaluations and judged risk. As participant experience increased, the magnitude of the relation between scenario liking and judged safety decreased. This moderating effect of experience was found in the analysis of the combined data from all four studies but was not found in any of the individual studies. However, the size of that effect was very small, making its practical relevance questionable. Care must be taken to not overemphasize the moderating effect of experience found in the pooled data, and further research is required to determine whether that observed effect is meaningful.

Do skiers' preferences influence their perception of risk?

We attempted to manipulate the attractiveness of the backcountry skiing scenarios while holding the risk level constant to test whether liking has a causal influence on perceived safety. Unfortunately, the failure of both manipulations to influence liking made it impossible to test causality. Nonetheless, those failed manipulations are relevant for the field of avalanche research and education. For the past two decades, a focus on erroneous decision heuristics (referred to as heuristic traps) has dominated examinations into the decision processes that result in avalanche accidents and fatalities (Johnson et al. 2020). One of the heuristic traps specified in that conceptual framework is the scarcity heuristic. It is based on the assumption that skiing untracked, fresh powder snow is so highly valued among backcountry skiers that they take greater risks to be the first to ski it (McCammon 2002, 2004). Evidence of the scarcity heuristic in previous research is conflicting. Furman, Shooter, and Schumann (2010) found evidence that an untracked slope was positively related to the likelihood of skiing that slope, whereas Marengo, Monaci, and Miceli (2017) found contradictory evidence that the presence of tracks increased the likelihood of skiing that slope. Our research found no evidence that the presence or absence of tracks influenced either liking or perceived safety. Another feature that avalanche experts believe influences the attractiveness of a ski slope and that they anecdotally associate with increased avalanche accidents is the sunny cloudless weather that can follow a night of snowfall, which skiers refer to as the highly prized *bluebird days* (Avalanche Canada 2016; Enright 2017; Morris 2016). Our research found no evidence that sunny versus cloudy skies influenced liking or perceived safety. These findings are applicable for the ongoing evaluation of the heuristic traps conceptual framework and its relevance for avalanche risk management strategies, tools and education (see Johnson et al. 2020). We recognize, however, that the task of judging avalanche risk may have focused participants' attention on risk to a greater degree than it would be in a natural situation. This increased focus on risk may have undermined any effect of their preference for the absence of tracks or for sunny blue skies. Future research could manipulate participants' focus on risk to test for an effect on liking.

Risk in extreme sports

The backcountry skiing context and those who perform the activity together provide an intriguing opportunity to study a real-world case of individuals who appear to like a dangerous activity. A prominent theoretical perspective on extreme sports contends that these activities are synonymous with risk and that participation is about risk-taking and the thrill it provides (Brymer, 2010). Moreover, measures of attitudes toward risk such as self-report questionnaires and lab-based tasks generally involve measuring affective disposition toward risk as the degree to which one likes taking risk (e.g., General Risk Propensity Scale (Zhang, Highhouse, and Nye 2019), Domain-Specific Risk Taking (DOSPERT) scale). A consequence of both that theoretical perspective on extreme sports and such measures of risk preference is that some people are deemed to like risk, defying the positive relation between liking and perceived safety. However, our data show that even people who like and perform a risky activity (at the category level) such as skiing in avalanche terrain exhibit a healthy positive relation between liking and perceived safety when judging specific instances of that activity (at the contextual level), rather than liking it because it is risky. This findings aligns with an alternative perspective on extreme sport participation that contends that people recognize the inherent risk of the activity, acknowledging the possibility of injury or death, but that they like safety and seek to maximize it when performing the activity (Brymer 2010).

Conclusion

Risk judgments and behavior decisions in the real world most often focus on specific instances of objects and activities rather than their categorical representations. Although prior research

demonstrated the positive relation between liking and perceived safety when judging categorical representations of stimuli, in the present research we have shown that that relation holds across multiple judgments of specific instances of an activity. Our findings contribute to understanding risk judgments and decision making as an interplay of cognitive and affective factors. That liking and perceived safety systematically vary, but that the positive relation between the two remains, when judging specific instances of an activity that a decision maker broadly likes and considers safe provides a dynamic picture of the degree to which cognitive and affective factors are intertwined.

Disclosure statement

No potential conflict of interest was provided by the author(s).

ORCID

Matthew B. Stephensen http://orcid.org/0000-0002-4562-3469
Torsten Martiny-Huenger http://orcid.org/0000-0003-3855-2890

Data availability statement

All data and analysis scripts are available at: https://doi.org/10.17605/OSF.IO/VA28N

References

Alhakami, A. S., and P. Slovic. 1994. "A Psychological Study of the Inverse Relationship between Perceived Risk and Perceived Benefit." *Risk Analysis : An Official Publication of the Society for Risk Analysis* 14 (6): 1085–1096. doi:10.1111/j.1539-6924.1994.tb00080.x.

Avalanche Canada 2016. Avalanche Rescue at Cherry Bowl: A Story that Needs to be Told. Retrieved from https://backcountryaccess.com/avalanche-rescue-at-cherry-bowl-a-story-that-needs-to-be-told/

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1): 1–48. doi:10.18637/jss.v067.i01.

Blanchette, I., and A. Richards. 2010. "The Influence of Affect on Higher Level Cognition: A Review of Research on Interpretation, Judgment, Decision Making and Reasoning." *Cognition & Emotion* 24 (4): 561–595. doi:10.18637/jss.v067.i01.

Brymer, E. 2010. "Risk Taking in Extreme Sports: A Phenomenological Perspective." *Annals of Leisure Research* 13 (1-2): 218–238. doi:10.1080/11745398.2010.9686845.

Colorado Avalanche Information Center. n.d. Forecasts. Retrieved from https://avalanche.state.co.us/

da Costa, B. R., and A. J. Sutton. 2019. "A Comparison of the Statistical Performance of Different Meta-Analysis Models for the Synthesis of Subgroup Effects from Randomized Clinical Trials." *BMC Medical Research Methodology* 19 (1): 198–198. doi:10.1186/s12874-019-0831-8.

Damasio, A. R. 1994. *Descartes' Error: Emotion, Reason and the Human Brain*. New York: Avon Books.

de Leeuw, J. R. 2015. "jsPsych: A JavaScript Library for Creating Behavioral Experiments in a Web Browser." *Behavior Research Methods* 47 (1): 1–12. doi:10.3758/s13428-014-0458-y.

Ebert, P. A. 2019. "Bayesian Reasoning in Avalanche Terrain: A Theoretical Investigation." *Journal of Adventure Education and Outdoor Learning* 19 (1): 84–95. doi:10.1080/14729679.2018.1508356.

Elliot, A. J., A. B. Eder, and E. Harmon-Jones. 2013. "Approach–Avoidance Motivation and Emotion: Convergence and Divergence." *Emotion Review* 5 (3): 308–311. doi:10.1177/1754073913477517.

Engeset, R. V., G. Pfuhl, M. Landrø, A. Mannberg, and A. Hetland. 2018. "Communicating Public Avalanche Warnings – What Works?" *Natural Hazards and Earth System Sciences* 18 (9): 2537–2559. doi:10.5194/nhess-18-2537-2018.

Enright, D. 2017. A Tragedy in the Japan Alps. Retrieved from https://www.evergreen-backcountry.com/a-tragedy-in-the-japan-alps/

Finucane, M. L., A. Alhakami, P. Slovic, and S. M. Johnson. 2000. "The Affect Heuristic in Judgments of Risks and Benefits." *Journal of Behavioral Decision Making* 13 (1): 1–17. %3C1::AID-BDM333%3E3.0.CO;2-S doi:10.1002/(SICI)1099-0771(200001/03)13:1.

Furman, N., W. Shooter, and S. Schumann. 2010. "The Roles of Heuristics, Avalanche Forecast, and Risk Propensity in the Decision Making of Backcountry Skiers." *Leisure Sciences* 32 (5): 453–469. doi:10.1080/01490400.2010.510967.

Gallatin National Forest Avalanche Center. n.d. Avalanche Forecast. Retrieved from https://www.mtavalanche.com/forecast

Hogarth, R. M., T. Lejarraga, and E. Soyer. 2015. "The Two Settings of Kind and Wicked Learning Environments." *Current Directions in Psychological Science* 24 (5): 379–385. doi:10.1177/0963721415591878.

Johnson, P. C. D., S. J. E. Barry, H. M. Ferguson, and P. Müller. 2015. "Power Analysis for Generalized Linear Mixed Models in Ecology and Evolution." *Methods in Ecology and Evolution* 6 (2): 133–142. doi:10.1111/2041-210X.12306.

Johnson, J.,. A. Mannberg, J. Hendrikx, A. Hetland, and M. Stephensen. 2020. "Rethinking the Heuristic Traps Paradigm in Avalanche Education: Past, Present and Future." *Cogent Social Sciences* 6 (1): 1807111. doi:10.1080/23311886.2020.1807111.

Kuznetsova, A., P. B. Brockhoff, and R. H. B. Christensen. 2017. "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software* 82 (13): 1–26. doi:10.18637/jss.v082.i13.

Lerner, J. S., Y. Li, P. Valdesolo, and K. S. Kassam. 2015. "Emotion and Decision Making." *Annual Review of Psychology* 66 (1): 799–823. doi:10.1146/annurev-psych-010213-115043.

Loewenstein, G. F., E. U. Weber, C. K. Hsee, and N. Welch. 2001. "Risk as Feelings." *Psychological Bulletin* 127 (2): 267–286. doi:10.1037/0033-2909.127.2.267.

Marengo, D., M. G. Monaci, and R. Miceli. 2017. "Winter Recreationists' Self-Reported Likelihood of Skiing Backcountry Slopes: Investigating the Role of Situational Factors, Personal Experiences with Avalanches and Sensation-Seeking." *Journal of Environmental Psychology* 49: 78–85. doi:10.1016/j.jenvp.2016.12.005.

McCammon, I. 2002. "*Evidence of Heuristic Traps in Recreational Avalanche Accidents*." Paper Presented at the Proceedings ISSW.

McCammon, I. 2004. "Heuristic Traps in Recreational Avalanche Accidents: Evidence and Implications." *Avalanche News* 68 (1): 42–50.

Medin, D. L. 1989. "Concepts and Conceptual Structure." *The American Psychologist* 44 (12): 1469–1481. doi:10.1037/0003-066x.44.12.1469.

Moors, A., and J. De Houwer. 2001. "Automatic Appraisal of Motivational Valence: Motivational Affective Priming and Simon Effects." *Cognition & Emotion* 15 (6): 749–766. doi:10.1080/02699930143000293.

Morris, S. 2016. Multiple avalanches in the Alps, two fatalities [weblog comment]. Retrieved from https://wepowder.com/en/forum/topic/219836

Niedermeier, M., H. Gatterer, E. Pocecco, A. Frühauf, M. Faulhaber, V. Menz, J. Burtscher, M. Posch, G. Ruedl, and M. Burtscher. 2019. "Mortality in Different Mountain Sports Activities Primarily Practiced in the Winter Season—a Narrative Review." *International Journal of Environmental Research and Public Health* 17 (1): 259. (. doi:10.3390/ijerph17010259.

R Core Team 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Schafer, J. L., and J. W. Graham. 2002. "Missing Data: Our View of the State of the Art." *Psychological Methods* 7 (2): 147–177. doi:10.1037/1082-989X.7.2.147.

Schauer, J. M., and L. V. Hedges. 2020. "Assessing Heterogeneity and Power in Replications of Psychological Experiments." *Psychological Bulletin* 146 (8): 701–719. doi:10.1037/bul0000232.

Schwarz, N. 2007. "Attitude Construction: Evaluation in Context." *Social Cognition* 25 (5): 638–656. doi:10.1521/soco.2007.25.5.638.

Schwarz, N. 2012. "Feelings-as-Information Theory." In *Handbook of Theories of Social Psychology*, edited by P. A. M. Van Lange, A. W. Kruglanski, & T. Higgins, Volume 1, 289–308. London: Sage Publications Ltd.

Schwarz, N., and G. L. Clore. 1983. "Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states." *Journal of Personality and Social Psychology* 45 (3): 513–523. doi:10.1037/0022-3514.45.3.513.

Schwarz, N., and G. L. Clore. 2007. "Feelings and phenomenal experiences." In *Social psychology: Handbook of basic principles*, edited by A. W. Kruglanski & E. T. Higgins, 385–407. New York, NY, US: Guilford Press.

Schwarz, N., and G. L. Clore. 1983. "Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states." *Journal of Personality and Social Psychology* 45 (3): 513–523. doi:10.1037/0022-3514.45.3.513.

Schwarz, N., and G. L. Clore. 2007. "Feelings and phenomenal experiences." In *Social psychology: Handbook of basic principles*, edited by A. W. Kruglanski & E. T. Higgins, 385–407. New York, NY, US: Guilford Press.

Sjöberg, L. 2006. "Will the Real Meaning of Affect Please Stand up?" *Journal of Risk Research* 9 (2): 101–108. doi:10.1080/13669870500446068.

Slovic, P., M. Finucane, E. Peters, and D. G. MacGregor. 2002. "The Affect Heuristic." In *Heuristics and Biases: The Psychology of Intuitive Judgment*, edited by D. Griffin, D. Kahneman, & T. Gilovich, 397–420. Cambridge: Cambridge University Press.

Slovic, P., M. L. Finucane, E. Peters, and D. G. MacGregor. 2004. "Risk as Analysis and Risk as Feelings: Some Thoughts about Affect, Reason, Risk, and Rationality." *Risk Analysis : An Official Publication of the Society for Risk Analysis* 24 (2): 311–322. doi:10.1111/j.0272-4332.2004.00433.x.

Tremblay, A., and J. Ransjin. 2020. *LMERConvenienceFunctions: Model Selection and Post-Hoc Analysis for (G)LMER Models*. R package version 3.0. https://CRAN.R-project.org/package=LMERConvenienceFunctions

Utah Avalanche Center. n.d. Forecasts. Retrieved from https://utahavalanchecenter.org/

Varsom n.d. *Avalanche Bulletins for Norway*. Retrieved from http://www.varsom.no/en/avalanche-bulletins/

Wardman, J. K. 2006. "Toward a Critical Discourse on Affect and Risk Perception." *Journal of Risk Research* 9 (2): 109–124. doi:10.1080/13669870500454773.

Westfall, J., D. A. Kenny, and C. M. Judd. 2014. "Statistical Power and Optimal Design in Experiments in Which Samples of Participants Respond to Samples of Stimuli." *Journal of Experimental Psychology. General* 143 (5): 2020–2045. doi:10.1037/xge0000014.

Wilson, R. S., and J. L. Arvai. 2006. "When Less is More: How Affect Influences Preferences When Comparing Low and High-Risk Options." *Journal of Risk Research* 9 (2): 165–178. doi:10.1080/13669870500419503.

Winkielman, P., R. B. Zajonc, and N. Schwarz. 1997. "Subliminal Affective Priming Resists Attributional Interventions." *Cognition & Emotion* 11 (4): 433–465. doi:10.1080/026999397379872.

Yee, E., and S. L. Thompson-Schill. 2016. "Putting Concepts into Context." *Psychonomic Bulletin & Review* 23 (4): 1015–1027. doi:10.3758/s13423-015-0948-7.

Zajonc, R. B. 1980. "Feeling and Thinking: Preferences Need No Inferences." *American Psychologist* 35 (2): 151–175. doi:10.1037/0003-066X.35.2.151.

Zhang, D. C., S. Highhouse, and C. D. Nye. 2019. "Development and Validation of the General Risk Propensity Scale (GRiPS)." *Journal of Behavioral Decision Making* 32 (2): 152–167. doi:10.1002/bdm.2102.

Zweifel, B., and P. Haegeli. 2014. "A Qualitative Analysis of Group Formation, Leadership and Decision Making in Recreation Groups Traveling in Avalanche Terrain." *Journal of Outdoor Recreation and Tourism* 5-6: 17–26. doi:10.1016/j.jort.2014.03.001.

Paper 3

Confidence in complex risk judgments: the roles of uncertainty, experience, and affect

Matthew B. Stephensen[1], Torsten Martiny-Huenger[1], and Christin Schulze[2,1]

[1]UiT The Arctic University of Norway, Tromsø, Norway

[2]Max Planck Institute for Human Development, Berlin, Germany

Open science statement: Data, analysis scripts, and study materials are available at https://osf.io/ba6eh/

Corresponding author: matthew.stephensen@uit.no; matthew.stephensen@gmail.com

Abstract

Disagreement persists about the origin of confidence and the internal signals that influence its formation. Using combined individual participant data from four studies (N = 181), we examined confidence in relation to the perceived source of uncertainty for a risk judgment, and explored the roles of domain-specific experience and affective evaluations in the formation of confidence. In each study, participants with domain-specific experience (backcountry skiers) performed complex risk judgments (judging avalanche risk) for multiple highly uncertain contexts (hypothetical avalanche terrain scenarios). We investigated whether more experienced participants could better recognize the inherent uncertainty of the decision environment, and if they did so with greater confidence. For complex tasks such as judging avalanche risk, experience should increase a person's understanding of the probabilistic, unpredictable nature of that environment. However, our findings suggests that participants of all levels of experience attributed uncertainty to their own judgment process rather than to the limitations and inherent uncertainty of the environment. We then examined whether participants' affective evaluations influenced their confidence in their risk judgments. Affective evaluations are understood to play a crucial orienting role in the risk judgment process. We found evidence of an interplay between affective and cognitive judgments in the formation of confidence. Participants were more confident when their affective evaluation matched their risk judgment, and less confident when there was a mismatch between the two. Our research illustrates a troubling limitation in the development of confidence with experience, and the potential (dis)advantageous effect of affective evaluations on confidence in certain contexts.

Keywords: Confidence, risk judgment, uncertainty, experience, affective evaluation

Confidence in complex risk judgments: the roles of uncertainty, experience, and affect

Doubt is unpleasant, but certainty is absurd

-Voltaire

Introduction

Judging risk involves evaluating the available evidence to make inferences about the likelihood of negative outcomes. Faced with uncertainty, people must have confidence in the accuracy of their risk judgments in order to act (Fitzgerald et al., 2017; Gill et al., 1998; Koriat & Goldsmith, 1996). Yet despite its importance, contention persists about the origin of confidence and the internal signals that influence its formation (Boldt et al. 2017; Meyniel et al., 2015; Petrusic & Baranski, 2003; Yeung & Summerfield, 2012). Although confidence has been extensively investigated using a variety of simple decision tasks, such as confidence in trivia knowledge, the applicability of this research to real-world judgements remains unclear (Dhami et al., 2004; Juslin et al., 2000; Koehler et al., 2002; Yeung & Summerfield, 2012). The complex, goal-oriented judgment tasks that typify human behavior often involve various sources of uncertainty and a multiplicity of cognitive processes and internal signals. In this article, we examine the formation of confidence in relation to the perceived source of uncertainty for a complex risk judgment task, and explore the roles of domain-specific experience and affective evaluations in that formation of confidence.

To capture the complexity and uncertainty of real-world decision environments, we asked backcountry skiers with varying levels of experience to judge the risk of snow avalanches in a series of uncertain hypothetical skiing scenarios modelled on the natural decision environment. Judging avalanche risk is a complex task for which no single cue or combination of cues allows perfect predictions. The reliability and validity of the available evidence are often limited, ambiguous, or conflicting. Consequently, there is always some uncertainty when assessing avalanche risk (Landrø et al., 2020). One might reasonably expect experienced backcountry skiers to confidently recognize that in many cases, avalanche risk in a complex environment is inherently probabilistic. By examining their confidence in relation to their risk judgments, we could ascertain whether they indeed recognized these situations as fundamentally uncertain with regard to the decision environment. We then investigated how domain-specific experience influenced that perception of uncertainty and their confidence judgments.

Uncertainty can take two qualitatively distinct forms based on its source. Internal uncertainty (i.e., epistemic uncertainty) is rooted in a decision maker's incomplete knowledge, skill, or information, while environmental uncertainty (i.e., aleatory uncertainty)

stems from the stochastic nature and inherent unpredictability of the environment (Fox & Ülkümen, 2011; Gillies, 2000; Kahneman & Tversky, 1982; Kozyreva & Hertwig, 2021; Løhre & Teigen, 2016; Peterson & Pitz, 1988; Teigen & Løhre, 2017; Ülkümen et al., 2016). Each form of uncertainty elicits a different type of probabilistic judgment. Internal uncertainty evokes an epistemic probabilistic judgment that involves reporting a subjective degree of belief in a predicted outcome, whereas environmental uncertainty evokes an aleatory probabilistic judgment that involves estimating the long-run frequency of an outcome given the external conditions (Fox & Ülkümen, 2011; Gillies, 2000; Kozyreva & Hertwig, 2021; Løhre & Teigen, 2016; Peterson & Pitz, 1988; Ülkümen et al., 2016). The perception of uncertainty and corresponding interpretation of the judgment task is highly subjective and contextual. People can disagree about the source of uncertainty for a particular judgment task, and a given individual can perceive different sources of uncertainty between specific instances of a category of judgment task (Fox & Ülkümen, 2011, Ülkümen et al., 2016).

Crucially, confidence in a risk judgment provides insight as to how a decision maker perceives the uncertainty of that judgment task. Because people make different probabilistic judgments in response to either form of uncertainty, the meaning of what is reported on a risk rating scale differs according to how the uncertainty of the judgment task is perceived (Fox & Ülkümen, 2011; Løhre & Teigen, 2016; Teigen & Løhre, 2017; Ülkümen et al., 2016). That meaning can be interpreted based on the retrospective confidence judgment. If the uncertainty of judging avalanche risk is attributed to a lack of knowledge and/or inadequacy of the judgment strategy, interpreting the task as an epistemic probabilistic judgment, then the reported risk value reflects a participant's subjective degree of belief in the safety/danger of the scenario (Fox & Ülkümen, 2011; Løhre & Teigen, 2016). Since confidence is equivalent to degree of belief for epistemic judgments, we would expect confidence to track that reported risk value: confidence should be highest at either limit of the response scale where the degree of belief in the safety/danger of the scenario is highest, and lowest at the midpoint of the response scale where the degree of belief is lowest. By contrast, if participants attribute uncertainty to the stochastic nature of the decision environment, interpreting the task as an aleatory probabilistic judgment, then reported risk reflects the long-run frequency of either outcome given the available evidence (Fox & Ülkümen, 2011; Løhre & Teigen, 2016). One can be more or less confident in any aleatory probabilistic estimate of the state of the world if the evidence supports that judgement. For example, when randomly drawing a card from a full deck, one can be highly confident that 25% of the time the suit of the card will be hearts. We could therefore expect equal confidence ratings across all values of judged safety. The

two forms of uncertainty evoke different probabilistic judgments and, correspondingly, different confidence judgments. By examining retrospective confidence reports in relation to what is reported on the risk rating scale, we can infer how participants perceive the uncertainty of the risk judgment. This is important because for complex tasks such as judging avalanche risk, people should ideally understand that the decision environment is probabilistic and that correctly judging the outcome is not exclusively dependent upon a decision maker's knowledge and judgment strategy.

In addition to examining confidence in relation to the perceived source of uncertainty, we explore the role of participants' backcountry skiing experience in the formation of confidence in avalanche risk judgments. Although experience is intuitively associated with expertise, the relationship between experience, confidence, and the degree to which a judgment corresponds with the true state of the world is not straightforward. As people gain experience, their judgment strategies tend to become more established resulting in greater confidence (Fischer & Budescu, 2005; Sanchez & Dunning, 2018, 2020). We might therefore expect experience with judging avalanche risk to correspond to reduced internal uncertainty and an increased understanding of the inherent environmental uncertainty. Unfortunately, greater experience and confidence do not guarantee that a person's judgment strategy represents a correct model of the world (Einhorn & Hogarth, 1978; Koriat, 2008; Hogarth et al, 2015). The acquisition of experience increases the conviction and tenacity with which a person holds to a judgment strategy more rapidly than it refines the content of that judgment strategy (DePaulo & Pfeifer, 1986; Sanchez & Dunning, 2018, 2020). We examine whether more experienced individuals are better able to recognize the inherent environmental uncertainty when judging avalanche risk, and if they have greater confidence in that recognition.

Beyond this examination of confidence in relation to experience and the perceived source of uncertainty, we also explore whether a backcountry skier's affective evaluation of a scenario influences their confidence in their judgment of the avalanche risk. Affective evaluations serve as a valuable and compelling orienting mechanism in the perception of risk (Blanchette & Richards, 2010; Finucane et al, 2000; Lerner et al., 2015; Lowenstein et al., 2001; Schwarz, 2012; Slovic, 1987; Slovic et al., 2002, 2004). Given that relation, it is reasonable to assume that affective evaluations might also influence the formation of confidence in risk judgments. That mood (i.e., incidental affect) has been found to influence confidence (Koellinger & Treffers, 2015; Kuvaas & Kaufmann, 2004; Massoni, 2014; Prinz et al., 2019; Sidi et al., 2018) is further reason to suspect that affective evaluations (i.e.,

integral affect) are an important internal signal during the formation of confidence in a complex risk judgment. We conceptualize affective evaluations as a generalized positive or negative response to the target of judgment, which we measure as scenario liking. Liking has a motivational component that is relevant for the risk judgment process: people are more likely to approach liked and avoid disliked stimuli (Chen & Bargh, 1999). Generally, liking positively relates to perceived safety. The more a backcountry skier likes (i.e., approach motivation) a scenario, the safer it is perceived, whereas the more they dislike (i.e., avoid motivation) a scenario, the more dangerous it is perceived (Alhakami & Slovic, 1994; Finucane et al., 2000; Slovic et al., 2002; Stephensen & Martiny-Huenger, 2021). However, despite this general relation, it is possible that an individual likes a scenario that is perceived as dangerous or dislikes a scenario perceived as safe. We explore whether the congruence (e.g., liking a scenario perceived as safe) and incongruence (e.g., liking a scenario perceived as dangerous) between affective evaluations and risk judgments influence the formation of confidence.

Methods

Participants. We obtained data from one-hundred and eighty-one individuals (59 women, 122 men, M_{age} = 31, *SD* = 11.18, range = 15-63) who participated in one of four independent studies. As explained below when reporting our methods of analysis, we combined the individual participant data from those four studies for pooled analysis. Participants in Studies 1 (N = 40) and 3 (N = 54) were backcountry skiers from Norway. Participants did not receive compensation for participating. Participants in Study 4 (N = 41) were backcountry skiers from the western USA who received compensation for participation. To obtain data from participants with a wide range of experience, participants in Study 2 (N = 46) were a mixed sample comprising backcountry skiers from Norway (37% of participants) who did not receive compensation for participating, and psychology students at a Norwegian university (63 % of participants) who received course credit for participating. The backcountry skier participants in each study had experience in performing similar risk judgments in preparation for and during actual ski tours in avalanche terrain. Arguably, the activity of backcountry skiing, the environment of avalanche terrain depicted in the scenarios, and the task of judging avalanche risk were familiar, meaningful, and important for those experienced participants.

We obtained informed consent for all participants in each study. All studies were conducted in accordance with the ethical research protocols of UiT – The Arctic University of Norway and the Norwegian Center for Research Data (NSD ref. numbers 54651 and 58839).

Data on judged safety, scenario liking, and participant experience from Studies 2 to 4 were previously published in the context of testing a hypothesized relation between scenario liking and judged safety (see Stephensen & Martiny-Huenger, 2021). However, confidence did not pertain to the focus of that investigation; we neither analyzed nor published data on confidence from any study in that article.

Materials. We conducted all four studies online using the jspsych programme (de Leeuw, 2015). We developed hypothetical scenarios each depicting a distinct backcountry ski tour on a snow-covered mountain slope (i.e., avalanche terrain, see Figure 1 for an example). Each scenario began with a description of the tour that included basic information on the terrain, elevation, average and maximum slope steepness, and the current weather. A photograph of the mountain slope was located below the tour information. Beneath the photograph was the forecasted regional avalanche danger level (5-category scale), the current avalanche problem(s) caused by the combination of weather and snow conditions, and the weather history. The avalanche danger scale and the avalanche problems are standardized tools used by the Norwegian Avalanche Warning Service (Varsom), the European Avalanche Warning Services (EAWS), the American Avalanche Association, and the National Avalanche Center in the USA. The photograph and textual content of each scenario provided objective information for judging the degree of avalanche risk (i.e., judged safety).

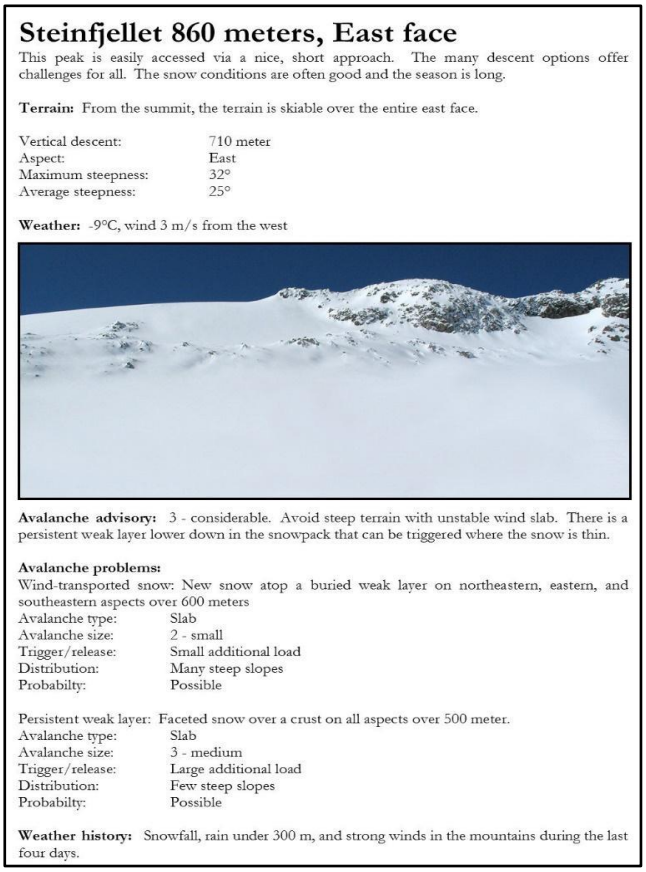
All the hypothetical scenarios represent realistic situations typically encountered in avalanche terrain. We designed each of the scenarios to have low predictability and high environmental uncertainty. None of the scenarios included indicators of either very low or very high risk of avalanche. The available information cues in the photo and textual content were ambiguous and/or conflicting. For example, although the danger level of 3-considerable (on a five-category scale), the weather history, and the prevailing avalanche problems in the scenario depicted in Figure 1 together indicate an increased probability of an avalanche, the terrain features, low slope angle, and the lack of any signs in the photograph indicative of the presence of the avalanche problems indicate a reduced probability of avalanche. It was not possible to ascertain that any scenario was safe or dangerous. Each scenario was marked by high environmental uncertainty.

Studies 1, 2, and 3 used the same six scenarios. For Study 4, we used six new scenarios. The content of all the scenarios such as the measurement units (e.g., Celsius and meters vs Fahrenheit and feet), geography, tour descriptions, and photographs were suited to the norms and familiar conditions of the participants. We based the content and format of the avalanche hazards information on historic avalanche forecasts from the Norwegian Avalanche

Warning Service (Varsom, n.d.) and from local US avalanche warning services (Colorado Avalanche Information Center, n.d.; Gallatin National Forest Avalanche Center, n.d.; Utah Avalanche Center, n.d.). All scenarios used in the four studies are available on the Open Science Framework at https://osf.io/ba6eh/.[1]

Figure 1

Example of the backcountry skiing scenarios judged by participants.



Steinfjellet 860 meters, East face

This peak is easily accessed via a nice, short approach. The many descent options offer challenges for all. The snow conditions are often good and the season is long.

Terrain: From the summit, the terrain is skiable over the entire east face.

Vertical descent:	710 meter
Aspect:	East
Maximum steepness:	32°
Average steepness:	25°

Weather: -9°C, wind 3 m/s from the west

Avalanche advisory: 3 - considerable. Avoid steep terrain with unstable wind slab. There is a persistent weak layer lower down in the snowpack that can be triggered where the snow is thin.

Avalanche problems:

Wind-transported snow: New snow atop a buried weak layer on northeastern, eastern, and southeastern aspects over 600 meters

Avalanche type:	Slab
Avalanche size:	2 - small
Trigger/release:	Small additional load
Distribution:	Many steep slopes
Probabilty:	Possible

Persistent weak layer: Faceted snow over a crust on all aspects over 500 meter.

Avalanche type:	Slab
Avalanche size:	3 - medium
Trigger/release:	Large additional load
Distribution:	Few steep slopes
Probabilty:	Possible

Weather history: Snowfall, rain under 300 m, and strong winds in the mountains during the last four days.

[1] We prepared two versions of each of the scenarios: one version included the clearly visible ski tracks on the mountain slope in the scenario photograph while the other version did not include ski tracks in the photograph (see Figure 1 for an example). As reported in Stephensen and Martiny-Huenger (2021), this within-subject manipulation (the presence vs. absence of ski tracks) was an attempt to manipulate scenario liking while holding constant the objective risk level of each scenario. However, this manipulation failed to influence judged safety, confidence, or scenario liking. We therefore do not investigate or discuss this manipulation further in this article.

Procedure. Participants accessed the online study via a web browser at a time and device of their choosing. Each scenario was presented on a unique webpage. The questions were presented immediately below each scenario on that same webpage. All the questions and response scales for each scenario were presented simultaneously and could be answered in any order. Participants in all four studies judged the safety of each scenario in terms of the avalanche risk, referred to as *judged safety*, and reported their retrospective confidence in judged safety, referred to as *confidence*. Participants in studies 2 to 4 also reported their affective evaluation of each scenario, referred to as *scenario liking*. For judged safety, participants answered three questions following the general statement "*Regarding the avalanche risk…*": 1) *"Is it safe to ski the slope in these conditions?"* 2) *"Is it dangerous to ski the slope in these conditions?"* (reverse-coded), and 3) *"Is the snowpack stable enough to ski this slope?"* On a 7-point scale labeled "*Not at all*" and "*Very much*" at the extreme points. For the degree of liking the scenario, following the general statement "*Regarding your desire to ski this slope, the slope appears…*" participants in Studies 2, 3, and 4 rated the scenario according to the following three adjectives: 1) *"Attractive", 2) "Uninteresting"* (reverse-coded), and *3) "Enjoyable"* on a 7-point scale labeled "*Not at all*" and "*Very much*" at the extreme points. We did not assess scenario liking in Study 1. For the degree of confidence in judged safety, participants answered three questions following the general statement "*Regarding your judgments of the avalanche risk…*": 1) *"Are you confident in your assessment of the avalanche risk?"*2) *"Do you doubt your judgment of the avalanche danger?"* (reverse-coded), and 3) *"Are you confident in your judgment of the safety of the snowpack?"* On a 7-point scale labeled "*Not at all*" and "*Very much*" at the extreme points. Participants were required to answer all questions for each scenario to proceed with the study. It was not possible to return to earlier scenarios. Participants in Study 1 judged each of the six available scenarios. Participants in Studies 2, 3, and 4 judged four scenarios from the sample of scenarios for their respective region. Selections were random for each participant. Upon completing the scenarios, participants reported their age and gender. Participants then reported their skiing ability, avalanche training, years of backcountry skiing experience, and average number of backcountry skiing days per season, each of which was measured on a 7-point scale.

Analysis. To synthesize the evidence across all studies and more accurately estimate parameters for any effect, we combined the individual participant data from each of the four studies for pooled analysis (da Costa & Sutton, 2019). The dataset comprises 804 scenario evaluations from the four studies. We calculated mean scores for confidence (three items,

Cronbach's alpha = .89, N = 804, M = 4.62, SD = 1.42), judged safety (three items, Cronbach's alpha = .92, N = 804, M = 3.18, SD = 1.44), and scenario liking (three items, Cronbach's alpha = .90, N = 564, M = 5.11, SD = 1.53) per participant and scenario. We calculated a mean experience score per participant from the measurements of skiing ability, avalanche training, years of backcountry skiing experience, and average number of backcountry skiing days per season (four items, Cronbach's alpha = .83, N = 181, M = 4.46, SD = 1.55). Three participants did not provide responses to all four measures of experience so we calculated mean experience scores using the available measures for those participants. We standardized the confidence, judged safety, scenario liking, and participant experience scores for all analyses.

We first analyzed whether an increase of experience corresponds to a similar increased ability to distinguish the source of uncertainty for a judgment task. We tested the variance of judged safety scores across all levels of participant experience for each scenario to determine if judged safety scores clustered among more experienced participants. We used linear regression analysis to determine whether experience predicts judged safety for each scenario. We used Breusch-Pagan tests of non-constant variance for that linear relation to determine if variance in judged safety was constant across participant experience levels for each scenario.

To investigate a possible U-shaped relation between judged safety and confidence, we used two-line regression analysis (Simonsohn, 2018) for all models predicting confidence. A U-shaped relation involves a sign change in the regression coefficient between confidence and the predictor variable. The point at which the regression coefficient changes between positive and negative is referred to as the breakpoint. We set the breakpoint at the midpoint of the judged safety reporting scale (safety judgment = 4 on a scale from 1 to 7). We split the data into two ranges according to the unstandardized judged safety score: one range in the direction of dangerous for judged safety scores < 4 and a second range in the direction of safe for judged safety scores ≥ 4. We fit identical models – as described in the next paragraph – for each of these ranges of judged safety.

We examined how confidence relates to the distribution of judged safety and any influence that experience and liking have on that relation. We estimated the effect of judged safety on confidence moderated by participant experience (*confidence ~ judged safety * participant experience + (1|participant) + (1|scenario) + (1|study)*) using the pooled data from all four studies. We then investigated the role of scenario liking in the formation of confidence using the combined data from Studies 2, 3, and 4. We excluded data from Study 1 in that analysis because it does not include the liking measure. We estimated the effect of

scenario liking on confidence when controlling for the effects of judged safety and experience (*confidence ~ judged safety * participant experience + scenario liking + (1|participant) + (1|scenario) + (1|study)*).

To estimate the parameters of those models, we fit Bayesian linear mixed models to evaluate the probability of the predictor variable(s) to predict the response variable confidence given the data. An advantage of Bayesian methods is the incorporation of (un)certainty caused by the amount of data, which is reflected in the width of the posterior distribution(s) of the model parameters (Nalborczyk et al., 2020). We used the *brms* package (Bürkner, 2017, 2018) in R (R Core Team, 2017) to generate posterior probability distributions to calculate median-based regression coefficients and 95% highest density intervals (HDI) to estimate parameters for each effect. We included varying intercepts for participants and scenarios in the models to account for by-subject and by-scenario variability. We also included varying intercepts for studies in the models to account for by-study clustering and the potential heterogeneity of effect parameter values between studies in the combined data. We defined a weakly informative prior distribution (Gelman et al., 2008, 2017) for the predictor judged safety (Normal(.2,1) for judged safety ≥ 4, Normal(-.2,1) for judged safety < 4) based on the assumption of a correlation between confidence and the extremeness of judged safety. We defined uninformed prior distributions (Normal(0,10)) for the possible effects of scenario liking and participant experience. We checked that Rhat values were less than 1.05 and greater than .9 to ensure chain convergence for all parameters of interest. Diagnostics indicated that all Rhat values were either 1.000 or 1.001 for all examined effects. The data and R script for data processing and analysis are publicly available on the Open Science Framework (https://osf.io/ba6eh/).

Results

Participants disagreed about the avalanche risk of each scenario. Figure 2 illustrates that judged safety scores per scenario were uniformly distributed across the response scale. Participants judged the risk heterogeneously, ranging from clearly safe to clearly dangerous, rather than clustering at the midpoint of the response scale. How then did confidence relate to that distribution of judged safety scores? As Figures 3 and 4 illustrate, confidence was lowest at the midpoint of the response scale and increased as judged safety became more extreme, be it in the direction of safe or dangerous. Table 1 presents the posterior distributions of all parameter estimates for the relation between judged safety and confidence. Participants did not report equal confidence across all values of judged safety. There was a U-shaped relation between judged safety and confidence.

Domain-specific experience did not reduce the variability of judged safety of each scenario. Linear regression and Breusch-Pagan tests (see the supplemental materials for results) indicate that experience did not predict judged safety and that the variance in judged safety remained constant across participant experience levels for all scenarios (except scenario 1 used in Studies 1 to 3). Although judged safety scores did not converge around the midpoint of the response scale at greater experience levels, experience magnified confidence in those judgments. Table 1 presents the posterior distributions of all parameter estimates for the influence of experience on confidence. As Figure 3 illustrates, higher participant experience predicted higher confidence, irrespective of judged safety values. Yet experience did not moderate the U-shaped relation between judged safety and confidence.

Scenario liking predicted confidence. Figure 4 illustrates that the effect of scenario liking on confidence was specific to whether a scenario was judged in the direction of safe or dangerous. Table 1 presents the posterior distributions of all parameter estimates for the relations between scenario liking and confidence. When a scenario was perceived as dangerous to some degree (judged safety < 4), confidence decreased with higher scenario liking. However, when a scenario was perceived as safe to some degree (judged safety ≥ 4), confidence increased with higher scenario liking. Liking a scenario perceived as safe or disliking a scenario perceived as dangerous increased confidence. By contrast, liking a scenario perceived as dangerous or disliking a scenario perceived as safe reduced confidence in the risk judgment.

Figure 2

Distribution of judged safety scores per scenario. Scenarios 1 to 6, which were used in Studies 1 to 3 (640 observations), show highly uniform distributions of judged safety scores. The distribution of judged safety scores for scenarios 7 to 12, which were only used in Study 4 (164 observations), are less uniform but nonetheless exhibit a large range of values indicating disparity in the perceived safety of each scenario.

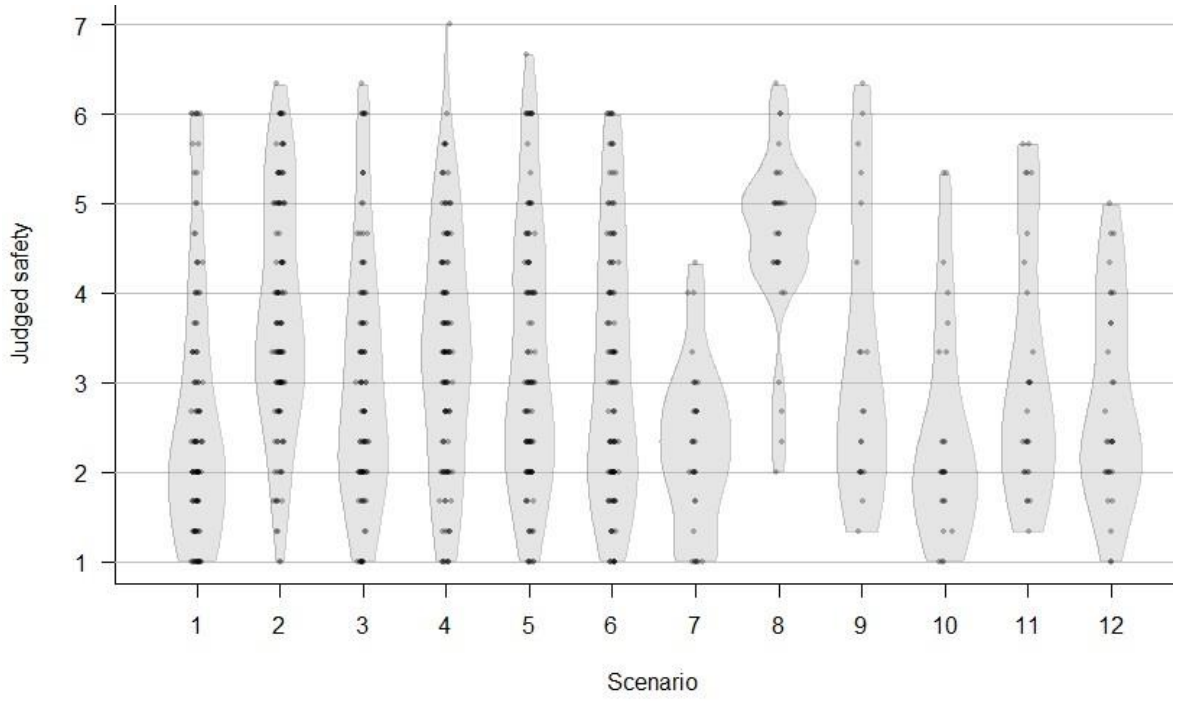


Figure 3

The effects of judged safety and participant experience on confidence (*confidence ~ judged safety * participant experience*) when a scenario is perceived as dangerous to some degree (plot on left, judged safety < 4) or perceived as safe to some degree (plot on right, judged safety ≥ 4). The effect of judged safety is plotted on the x-axis. The effect of participant experience is plotted as a solid line for -1 SD participant experience, a dashed line for mean participant experience, and a dotted line for +1 SD participant experience.

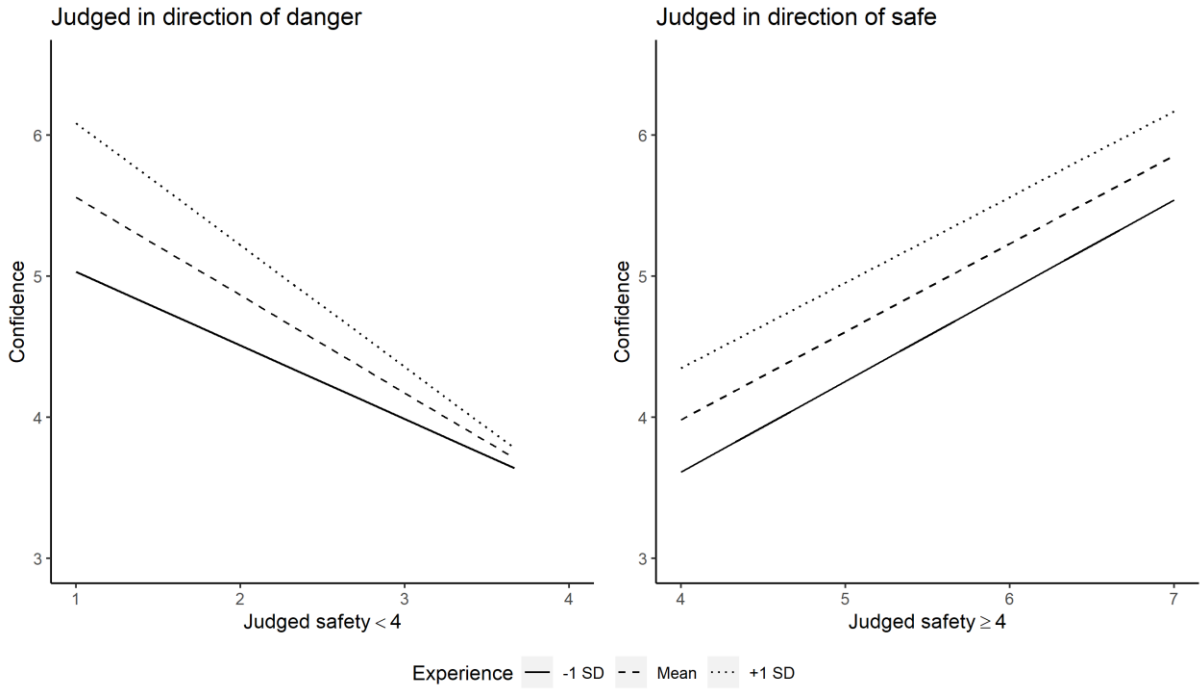


Figure 4

The effects of judged safety, scenario liking, and participant experience on confidence (*confidence ~ judged safety * participant experience + scenario liking*) when a scenario is perceived as dangerous to some degree (plot on left, judged safety < 4) or perceived as safe to some degree (plot on right, judged safety ≥ 4). The effect of judged safety is plotted on the x-axis. The effect of participant experience is plotted as a solid line for -1 SD participant experience, a dashed line for mean participant experience, and a dotted line for +1 SD participant experience. The effect of scenario liking is faceted into -1 SD scenario liking in the left column of each plot, mean scenario liking in the middle column of each plot, and +1 SD scenario liking in the right column of each plot.

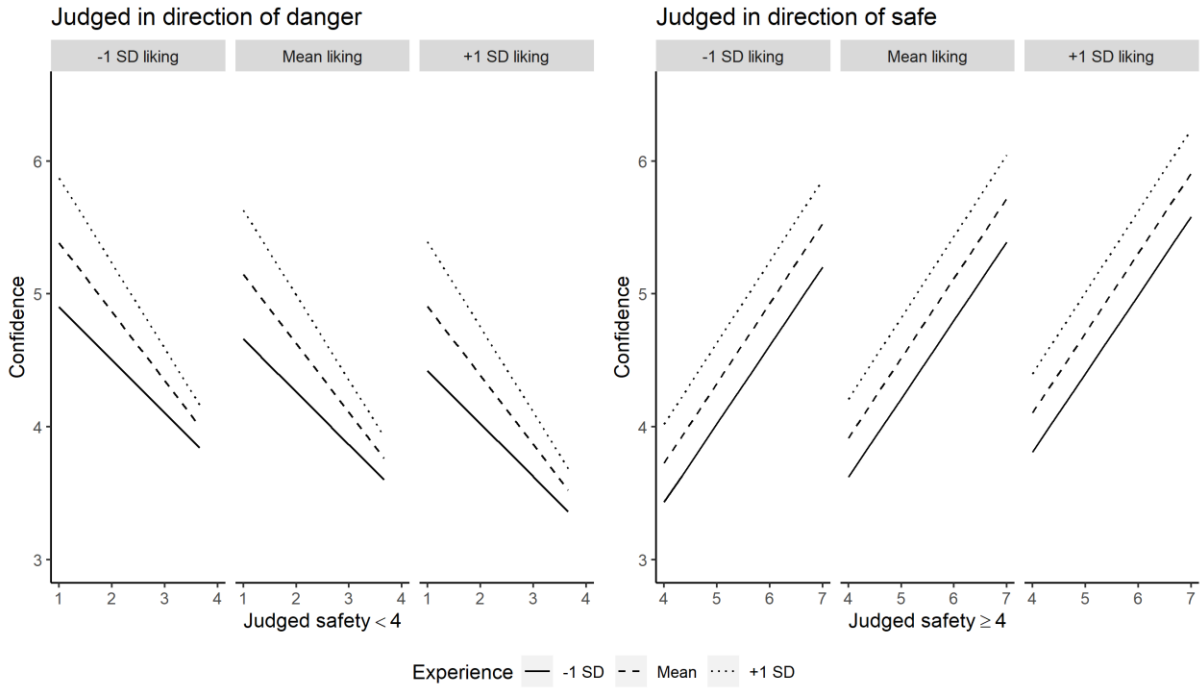


Table 1

Posterior parameter distributions of the effects predicting confidence in the judgment that a scenario is dangerous to some degree (judged safety < 4) or is safe to some degree (judged safety ≥ 4).

	Judged safety < 4		Judged safety ≥ 4	
Effect	*b* (*SE*)	95% HDI	*b* (*SE*)	95% HDI
*Confidence ~ judged safety * experience*				
Judged safety	-.69 (.06)	-.82, -.57	.63 (.08)	.47, .79
Experience	.10 (.08)	-.05, .25	.28 (.11)	.06, .50
Judged safety * experience	-.17 (.06)	-.28, -.06	-.02 (.08)	-.18, .14
*Confidence ~ judged safety * experience + scenario liking*				
Judged safety	-.55 (.08)	-.71, -.39	.61 (.10)	.42, .80
Experience	.15 (.09)	-.02, .33	.22 (.13)	-.04, .47
Scenario liking	-.16 (.05)	-.26, -.05	.19 (.08)	.04, .34
Judged safety * experience	-.12 (.07)	-.25, .02	.01 (.09)	-.17, .19

Discussion

Experience with a judgment task ideally should reduce internal, epistemic uncertainty by increasing the accuracy of a person's judgment theories and practices. In a highly complex decision environment, experience should also increase a person's understanding of the probabilistic nature of that environment. More experienced backcountry skiers should be better able to recognize when the complexity of the environment is so great that it exceeds their ability to reliably determine whether the conditions are safe or dangerous. Yet this is not what we found. Confidence in avalanche risk judgments increased with domain-specific experience, echoing previous findings on the development of confidence (Fischer & Budescu, 2005; Sanchez & Dunning, 2018, 2020). Yet that greater confidence did not correspond to an increased ability to recognize the inherent uncertainty of the scenarios. More experienced participants were not better able to understand that the safety or danger of each scenario could not be established. Our pattern of finding suggests that participants of all levels of experience attributed uncertainty to their own judgment process rather than to the limitations and inherent uncertainty of the environment.

That people fail to recognize the unpredictability of the decision environment and instead attribute uncertainty to their own incomplete knowledge, competency, or information is problematic from a policy perspective. Ideally, we want backcountry skiers in avalanche terrain – and decision makers in other fields performing similarly complex risk judgments – to understand when the safety or danger of a situation cannot be established because of the inherent uncertainty of the decision environment. We do not want people mistakenly assuming that safety or danger can always be determined or that they have the expertise to outthink the information limits and uncertainty of the environment. Unfortunately, the experience of avalanche terrain can lead many backcountry skiers to make those mistaken assumptions. Avalanche terrain is a decision environment in which poor decisions are infrequently marked by valid feedback (Ebert, 2019; Johnson et al., 2020; Zweifel & Haegeli, 2014). Valid feedback is essential for the experiential learning and calibration process (Fischer & Budescu, 2005; Sanchez & Dunning, 2018, 2020). The absence of valid performance feedback makes it problematic for decision makers to evaluate their performance and refine their judgment strategy. This obstructs skill development and magnifies the likelihood of overconfidence in erroneous decision making practices (Hogarth et al., 2015). Although confidence increased with domain-specific experience, even the most experienced participants did not confidently recognize that the conflicting, ambiguous evidence made it impossible to establish the safety or danger of any scenario. In this regard, experience did not correspond to expertise.

We recognize that the design of the scenarios and the studies might offer alternative explanations for the high variance in judged safety across all levels of experience and the perceived source of uncertainty. The use of exclusively uncertain scenarios may have been ecologically invalid for an experienced participant sample (Dhami et al, 2004). In the real world, experienced participants will have previously encountered situations that are objectively safe or dangerous among other situations that are uncertain. The distribution of reported judged safety and the confidence in those judgments per participant may partly reflect the probabilistic distribution (i.e., base rate) of safety, danger, and uncertainty encountered in their own real-world experience. However, had ecological validity been crucial, we would expect less experienced participants to have shown less variance in their risk judgments because their prior assumptions about base rates of safe, uncertain, and dangerous conditions are weaker. That they did not suggests that the lack of ecological validity was not a problem.

Another possible limitation concerns the interpretation of confidence to infer the perceived source of uncertainty. By asking participants about their confidence, we may have prompted them to assess their degree of belief in their risk judgment rather than the likelihood of their risk judgment being correct. Previous research (Fox & Ülkümen, 2011; Løhre & Teigen, 2016; Ülkümen et al., 2016) has shown that people's choice of language when communicating uncertainty differs as function of where they locate the source of uncertainty. Aleatory or environmental uncertainty is expressed in likelihood statements (e.g., "I'd say there is a 90% chance," or "I think there is a high probability") whereas epistemic uncertainty is expressed in confidence statements (e.g., "I am 80% sure," or "I am reasonably certain"). Because the perception of uncertainty influences the choice of language, one might conversely assume that the language used to formulate a judgment task could influence the perception of uncertainty. By using the language of confidence when asking participants to retrospectively judge the accuracy of their risk judgments, we may have made epistemic uncertainty more salient. However, the relation between language and perceived uncertainty was found when the choice of language pertained to judging a possible outcome or state (e.g., "I am __% sure that Germany will win the FIFA World Cup" versus "I think there is a __% chance that Germany will win the FIFA World Cup"). That relation has not been demonstrated (or, to the best of our knowledge, investigated) when the choice of language pertains to retrospective judgments of performance (i.e., metacognitive judgments). Judging performance accuracy is different from judging a possible outcome.

The final question we examined concerned whether affective evaluations influenced confidence. Our research found evidence of an interplay between affective and cognitive judgments in the formation of confidence. Judged safety and scenario liking each served as a cue for the formation of confidence, and their relative configuration introduced additional variability into confidence reports. Agreement between the cognitive and affective evaluations (liking scenarios perceived as safe, disliking scenarios perceived as dangerous) magnified confidence whereas disagreement between the two (liking scenarios perceived as dangerous, disliking scenarios perceived as safe) decreased confidence. The influence of the (in)congruity between cognitive and affective evaluations appears beneficial under uncertainty. Confidence functions as a control process for whether a decision is translated into action (Fitzgerald et al., 2017; Gill et al., 1998; Koriat & Goldsmith, 1996). Cognitive and affective evaluations appear to check and balance each other to prevent overconfidence and mistaken actions. A disagreement between cognitive and affective evaluations would suggest that either evaluation overlooked or misjudged a valid, relevant cue(s) or mistakenly relied on

an invalid, irrelevant cue(s). If both cognitive and affective evaluation processes are well calibrated for a judgment task, reduced confidence when there is disagreement between the two would be highly advantageous because it would decrease the likelihood of acting on mistaken evaluations. However, if an affective evaluation is based on irrelevant features, a negative affective response could diminish confidence in an objectively correct judgment of danger.

In conclusion, our research provides insight into some of the multiple internal signals that influence the formation of confidence in complex risk judgments under uncertainty. Specifically, our findings illustrate a troubling limitation of the development of confidence with experience and the potential (dis)advantageous effect of affective evaluations on confidence in certain contexts.

References

Alhakami, A. S., & Slovic, P. (1994). A Psychological Study of the Inverse Relationship Between Perceived Risk and Perceived Benefit. *Risk Analysis, 14*(6), 1085-1096. https://doi.org/10.1111/j.1539-6924.1994.tb00080.x

Blanchette, I., & Richards, A. (2010). The influence of affect on higher level cognition: A review of research on interpretation, judgment, decision making and reasoning. *Cognition and Emotion, 24*(4), 561-595. https://doi.org/10.18637/jss.v067.i01

Boldt, A., de Gardelle, V., & Yeung, N. (2017). The impact of evidence reliability on sensitivity and bias in decision confidence. *Journal of Experimental Psychology: Human Perception and Performance, 43*(8), 1520-1531. https://doi.org/10.1037/xhp0000404

Bürkner, P-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software, 80,* 1-28. https://doi.org/10.18637/jss.v080.i01

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. Personality and Social Psychology Bulletin, 25, 215–224. https://doi.org/10.1177/0146167299025002007

Colorado Avalanche Information Center. (n.d.). *Forecasts*. Retrieved from https://avalanche.state.co.us/

da Costa, B. R., & Sutton, A. J. (2019). A comparison of the statistical performance of different meta-analysis models for the synthesis of subgroup effects from randomized clinical trials. *BMC medical research methodology, 19*(1), 198-198. https://doi.org/10.1186/s12874-019-0831-8

Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The Role of Representative Design in an Ecological Approach to Cognition. *Psychological Bulletin, 130*(6), 959-988. https://doi.org/10.1037/0033-2909.130.6.959

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods, 47*, 1-12. https://doi.org/10.3758/s13428-014-0458-y

DePaulo, B. M., & Pfeifer, R. L. (1986). On-the-job experience and skill at detecting deception. *Journal of Applied Social Psychology, 16*(3), 249-267. https://doi.org/10.1111/j.1559-1816.1986.tb01138.x

Ebert, P. A. (2019). Bayesian reasoning in avalanche terrain: a theoretical investigation. *Journal of Adventure Education and Outdoor Learning, 19*(1), 84-95. https://doi.org/10.1080/14729679.2018.1508356

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review*, *85*, 395–416. https://doi.org/10.1037/0033-295X.85.5.395

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making, 13*(1), 1-17. https://doi.org/10.1002/(SICI)1099-0771(200001/03)13:1%3C1::AID-BDM333%3E3.0.CO;2-S

Fischer, I., & Budescu, D. V. (2005). When do those who know more also know more about how much they know? The development of confidence and performance in categorical decision tasks. *Organizational Behavior and Human Decision Processes, 98*(1), 39-53. https://doi.org/10.1016/j.obhdp.2005.04.003

Fitzgerald, L. M., Arvaneh, M., & Dockree, P. M. (2017). Domain-specific and domain-general processes underlying metacognitive judgments. *Consciousness and Cognition, 49*(Supplement C), 264-277. https://doi.org/10.1016/j.concog.2017.01.011

Fox, C. R., & Ülkümen, G. (2011). Distinguishing Two Dimensions of Uncertainty. In W. Brun, G. Keren, G. Kirkebøen, & H. Montgomery (Eds.), *Perspectives on Thinking, Judging, and Decision Making: A Tribute to Karl Halvor Teigen*. Oslo: Universitetsforlaget.

Gallatin National Forest Avalanche Center. (n.d.). *Avalanche Forecast*. Retrieved from https://www.mtavalanche.com/forecast

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. The Annals of Applied Statistics, 2(4), 1360-1383, 1324. https://doi.org/10.1214/08-AOAS191

Gelman, A., Simpson, D., & Betancourt, M. (2017). The Prior Can Often Only Be Understood in the Context of the Likelihood. *Entropy*, *19*(10), 555. https://doi.org/10.3390/e19100555

Gill, M. J., Swann Jr, W. B., & Silvera, D. H. (1998). On the genesis of confidence. *Journal of Personality and Social Psychology*, *75*(5), 1101-1114. https://doi.org/10.1037/0022-3514.75.5.1101

Gillies, D. (2000). *Philosophical theories of probability*. London: Routledge.

Hogarth, R. M., Lejarraga, T., & Soyer, E. (2015). The Two Settings of Kind and Wicked Learning Environments. *Current Directions in Psychological Science, 24*(5), 379-385. https://doi.org/10.1177/0963721415591878

Johnson, J., Mannberg, A., Hendrikx, J., Hetland, A., & Stephensen, M. (2020). Rethinking the Heuristic Traps Paradigm in Avalanche Education: Past, Present and Future. *Cogent Social Sciences, 6*(1). https://doi.org/10.1080/23311886.2020.1807111

Juslin, P., Winman, A., & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard–easy effect. *Psychological Review, 107*(2), 384-396. https://doi.org/10.1037/0033-295X.107.2.384

Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition, 11*(2), 143-157. https://doi.org/10.1016/0010-0277(82)90023-3

Koehler, D. J., Brenner, L., & Griffin, D. (2002). The Calibration of Expert Judgment: Heuristics and Biases Beyond the Laboratory. In D. Griffin, D. Kahneman, & T. Gilovich (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 686-715). Cambridge: Cambridge University Press.

Koellinger, P., & Treffers, T. (2015). Joy Leads to Overconfidence, and a Simple Countermeasure. *PLoS ONE, 10*(12), e0143263. https://doi.org/10.1371/journal.pone.0143263

Koriat, A. (2008). When confidence in a choice is independent of which choice is made. *Psychonomic Bulletin & Review, 15*, 997–1001. https://doi.org/10.3758/PBR.15.5.997

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*(3), 490-517. https://doi.org/10.1037/0033-295X.103.3.490

Kozyreva, A., & Hertwig, R. (2021). The interpretation of uncertainty in ecological rationality. *Synthese, 198*(2), 1517-1547. https://doi.org/10.1007/s11229-019-02140-w

Kuvaas, B., & Kaufmann, G. (2004). Impact of mood, framing, and need for cognition on decision makers' recall and confidence. *Journal of Behavioral Decision Making, 17*(1), 59-74. https://doi.org/10.1002/bdm.461

Landrø, M., Pfuhl, G., Engeset, R., Jackson, M., & Hetland, A. (2020). Avalanche decision-making frameworks: Classification and description of underlying factors. Cold Regions Science and Technology, 169, 102903. https://doi.org/10.1016/j.coldregions.2019.102903

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and Decision Making. Annual Review of Psychology, 66(1), 799-823. https://doi.org/10.1146/annurev-psych-010213-115043

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin, 127*(2), 267-286. https://doi.org/10.1037/0033-2909.127.2.267

Løhre, E., & Teigen, K. H. (2016). There is a 60% probability, but I am 70% certain: communicative consequences of external and internal expressions of uncertainty. *Thinking & Reasoning, 22*(4), 369-396. https://doi.org/10.1080/13546783.2015.1069758

Massoni, S. (2014). Emotion as a boost to metacognition: How worry enhances the quality of confidence. *Consciousness and Cognition, 29*, 189-198. https://doi.org/10.1016/j.concog.2014.08.006

Meyniel, F., Schlunegger, D., & Dehaene, S. (2015). The Sense of Confidence during Probabilistic Learning: A Normative Account. *PLOS Computational Biology, 11*(6), e1004305. https://doi.org/10.1371/journal.pcbi.1004305

Nalborczyk, L., Batailler, C., Loevenbruck, H., Vilain, A., & Bürkner, P. (2019, May 14). An Introduction to Bayesian Multilevel Models Using brms: A Case Study of Gender Effects on Vowel Variability in Standard Indonesian. https://doi.org/10.1044/2018_JSLHR-S-18-0006

Peterson, D. K., & Pitz, G. F. (1988). Confidence, uncertainty, and the use of information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(1), 85–92. https://doi.org/10.1037/0278-7393.14.1.85

Petrusic, W.M., & Baranski, J.V. (2003). Judging confidence influences decision processing in comparative judgments. *Psychonomic Bulletin & Review, 10***,** 177–183. https://doi.org/10.3758/BF03196482

Prinz, A., Bergmann, V., & Wittwer, J. (2019). Happy but overconfident: positive affect leads to inaccurate metacomprehension. Cognition and Emotion, 33(3), 606-615. https://doi.org/10.1080/02699931.2018.1472553

R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Sanchez, C., & Dunning, D. (2018). Overconfidence among beginners: Is a little learning a dangerous thing? *Journal of Personality and Social Psychology, 114*(1), 10-28. https://doi.org/10.1037/pspa0000102

Sanchez, C., & Dunning, D. (2020). Decision fluency and overconfidence among beginners. *Decision, 7*(3), 225-237. https://doi.org/10.1037/dec0000122

Schwarz, N. (2012). Feelings-as-Information Theory. In P. A. M. Van Lange, A. W. Kruglanski, & T. Higgins (Eds.), *Handbook of Theories of Social Psychology: Volume 1* (pp. 289-308). London: Sage Publications Ltd.

Sidi, Y., Ackerman, R., & Erez, A. (2018). Feeling happy and (over)confident: the role of positive affect in metacognitive processes. Cognition and Emotion, 32(4), 876-884. https://doi.org/10.1080/02699931.2017.1346501

Simonsohn, U. (2018). Two Lines: A Valid Alternative to the Invalid Testing of U-Shaped Relationships With Quadratic Regressions. *Advances in Methods and Practices in Psychological Science, 1*(4), 538-555. https://doi.org/10.1177/2515245918805755

Slovic, P. (1987). Perception of risk. *Science, 236*(4799), 280. https://doi.org/10.1126/science.3563507

Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). The Affect Heuristic. In D. Griffin, D. Kahneman, & T. Gilovich (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 397-420). Cambridge: Cambridge University Press.

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2004). Risk as Analysis and Risk as Feelings: Some Thoughts about Affect, Reason, Risk, and Rationality. *Risk Analysis, 24*(2), 311-322. https://doi.org/10.1111/j.0272-4332.2004.00433.x

Stephensen, M. B., & Martiny-Huenger, T. (2021). Liking and perceived safety across judgments of distinct instances of a category of activity. *Journal of Risk Research*, 1-17. https://doi.org/10.1080/13669877.2021.1905693

Teigen, K. H., & Løhre, E. (2017). Expressing (un)certainty in no uncertain terms: reply to Fox and Ülkümen. *Thinking & Reasoning, 23*(4), 492-496. https://doi.org/10.1080/13546783.2017.1314965

Ülkümen, G., Fox, C. R., & Malle, B. F. (2016). Two dimensions of subjective uncertainty: Clues from natural language. *Journal of Experimental Psychology: General, 145*(10), 1280-1297. https://doi.org10.1037/xge0000202

Utah Avalanche Center. (n.d.). *Forecasts*. Retrieved from https://utahavalanchecenter.org/

Varsom. (n.d.). *Avalanche Bulletins for Norway*. Retrieved from http://www.varsom.no/en/avalanche-bulletins/

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1594), 1310-1321. https://doi.org10.1098/rstb.2011.0416

Zweifel, B., & Haegeli, P. (2014). A qualitative analysis of group formation, leadership and decision making in recreation groups traveling in avalanche terrain. *Journal of Outdoor Recreation and Tourism, 5-6*, 17-26. https://doi.org/10.1016/j.jort.2014.03.001

Supplemental Materials

Table 1

Results of linear regression analysis of the relation between experience and judged safety per scenario used in Studies 1 to 4, and the results Breusch-Pagan tests of non-constant variance for that linear relation. The null hypothesis for the Breusch-Pagan test assumes homoscedasticity of variance.

Scenario	Judged safety ~ experience			Breusch-Pagan test (df = 1)	
	$b_{experience}$ (*SE*)	*95% CI*	*p*-value	χ^2	*p*-value
1	- .15 (.06)	-.26, -.04	.007	3.69	.055
2	.01 (.05)	-.10, .11	.864	1.02	.313
3	- .08 (.07)	-.22, .06	.249	.17	.684
4	- .01 (.05)	-.11, .09	.858	.64	.424
5	- .02 (.06)	-.15, .11	.761	.01	.927
6	.00 (.06)	-12, .11	.950	.00	.965
7	- .03 (.14)	-.30, .25	.835	.37	.544
8	- .06 (.12)	-.31, .18	.603	.63	.426
9	- .32 (.22)	-78, .13	.155	.89	.346
10	- .12 (.18)	-.51, .26	.505	.05	.817
11	.06 (.20)	-.34, .47	.746	.38	.535
12	.02 (.14)	-.26, .31	.859	.16	.688

Appendix 1

This popular science article based on Paper 1 was published in the spring 2021 volume of *The Avalanche Journal*.[2] The article was re-published in the June volume of the *New Zealand Avalanche Dispatch*.[3]

Should We Judge Danger or Safety in Avalanche Terrain?

Matthew B Stephensen
Markus Landrø
Jordy Hendrikx

Mindy and Kelsey hike up the valley, weaving their way through thinning forest and steepening terrain to the top of a small rise where they stop. They are entering avalanche terrain and it is time to decide if they should continue on their planned route. Ahead of them is a long, steep climb up a broad face to reach the more gradual ridgeline that they intend to follow to the summit. They dig a snow pit but do not find any sign of the persistent weak layer mentioned in the regional avalanche forecast. The snowfall has been light but steady and the winds variable over the past 48 hours. Although no cornices are visible on the ridgeline, spindrift indicates the wind is starting to pick up. They have not seen any obvious avalanche clues. They stand there, pondering the uncertainty of the conditions.

Mindy breaks the silence: “It looks good. I don’t believe it’s dangerous,” she remarks. “I think we should continue.”

Kelsey wrinkles her brow: “Really? It doesn’t look safe to me,” she counters. “We should turn back.”

---

[2] Stephensen, M.B., Landrø, M., & Hendrikx, J. (2021). Should We Judge Danger or Safety in Avalanche Terrain? *The Avalanche Journal, 126*, 18-19.
[3] Stephensen, M.B., Landrø, M., & Hendrikx, J. (2021). Should We Judge Danger or Safety in Avalanche Terrain? *New Zealand Avalanche Dispatch*, *June*, 47-50.

Why might two recreationists with similar training, competency, and experience make opposite decisions when judging the same evidence about the conditions? To try to answer that question, we must consider the cognitive mechanisms involved in the decision making process.

Question Framing

When we judge risk, we are judging the attribute of an object, action, or situation. Attributes are commonly understood in terms of their multiple dimensions. For example, the attribute 'speed' is often understood in terms of two dimensions: fast and slow. Those two dimensions are like two sides of a coin, distinct but indivisible. They provide opposing but complementary perspectives.

We tend to focus on a single dimension when making a judgment. For example, when judging speed, we commonly ask "Is it *fast*?" or, alternatively, "Is it *slow*?" rather than formulating a judgment using both dimensions. It is a natural process of language and thought to frame judgments with only one qualitative dimension of the judged attribute.

Similarly, when touring in avalanche terrain, we might also use a single qualitative dimension to frame avalanche risk judgements such as "How *safe* are the conditions?" or "How *dangerous* are the conditions?"

What we wanted to know is: does the choice of frame have an effect on perceived risk and behavioural decisions? If so, might we strategically employ that frame to increase the likelihood of more cautious, conservative judgments and decisions in avalanche terrain?

Our research found that frames influence perceived avalanche risk and behaviour intention. In a series of studies, we examined how backcountry skiers judged hypothetical scenarios of skiing in avalanche terrain (presented in the form of a photo and basic regional avalanche advisory information) when asked to judge *safety* or *danger*. We found that risk judgments framed in terms of safety ("*How safe is it?"*) resulted in more cautious, conservative judgments and a lower likelihood of skiing than judgments framed in terms of danger ("*How dangerous is it?"*).

This happens because the frames "safe" or "dangerous" direct the decision maker's attention during the judgement process. Judging "*How safe is it?*" defines *safe* as the reference point for the risk judgment. This focuses attention on finding and evaluating evidence of safety. Under conditions of uncertainty when there is no definitive indication of safety, such as in our example with Mindy and Kelsey, safety is judged as lower due to the lack of supporting evidence. Lower

safety implies the unspecified opposite dimension—higher danger—resulting in a lower likelihood of deciding to ski.

Conversely, when judging the danger (*How dangerous is it?*) of the same scenarios, danger is judged to be lower (and consequently safety is perceived to be higher) because of the lack of definitive evidence of danger, resulting in a higher likelihood of deciding to ski.

By asking backcountry travellers to judge how safe the conditions are, we exploit the lack of definitive evidence of safety to actually promote more cautious judgements and behaviour. Failing to find evidence of danger should not be considered an indication of safety. Yet failing to find evidence of safety must be considered an indication that it is not safe.

Strategically Framing Risk Perception in Avalanche Terrain

Decision making in avalanche terrain is seldom free of uncertainty. How we formulate risk judgments can have a real impact on how people perceive risk and, ultimately, when and how they decide to act. How then might we harness the power of framing in the avalanche industry?

The use of framing to promote specific judgments and decisions is an established practice in fields such as media and marketing. We can similarly employ framing to promote safer risk perceptions and behaviour in avalanche terrain. Guides, avalanche warning services, and avalanche safety educators have substantial control over the phrasing of questions about the risks they assess for a backcountry trip. Framing can be systematically applied to numerous risk judgments and decisions, whether it be deciding to ski a specific line or deciding to open or close specific terrain.

Communication between members of a group travelling in avalanche terrain could be positively impacted by increased awareness of the framing effect. How information and questions are framed could influence other group members' perceptions of the current risk and the decisions made or communicated between members of the group. For example, when a guide or group leader notices changes in the conditions, they can advantageously frame their question to the group to focus attention on those changes in relation to the basis for any earlier judgments of safety.

Let's say that the basis for the decision to ascend a slope is that the old snowpack is stable with fresh, non-wind loaded powder snow on top. After some climbing, the snow surface shows signs of wind effect. Focusing on establishing safety forces the group to reassess the conditions

relative to the previous evidence of safety (non-wind affected snow) and the possibility that conditions have changed (evidence of wind slabs). The group must evaluate if the evidence previously indicating safety is no longer present or if new evidence of safety is available. The group must therefore reconsider its arguments and possibly change its decision.

Asking "How safe is this slope?" increases attention paid towards evidence of safety—not just the absence of signs of danger—making the group more critically aware of any changes in conditions while guiding the decision toward a more conservative, transparent, and possibly safer outcome.

Framing risk judgments alone is insufficient to ensure safer behaviour among all backcountry recreationists. It is no substitute for the training, knowledge, and experience to understand and apply information about the conditions. Nonetheless, adopting a strategy for framing risk judgments can increase the likelihood of more cautious, conservative behaviour.

There is often so much uncertainty when making decisions in avalanche terrain that we must utilize any tool or method that can help, even if just a little. Critically, one wants to minimize the chance that a "*go*" decision is made under objectively "*no-go*" conditions. If the way a question is framed influences the decision of whether to ride or not, then avalanche risk management strategies, tools, and education should recognize and account for this effect and incorporate framing risk judgments into routine practices to reduce the potential for avalanche accidents.

The next time there is uncertainty about the current avalanche conditions, stop and think about how the way you frame the question could influence your users' decisions and think about how you might be able to harness framing to provide that extra margin of safety in times of uncertainty.

A peer-reviewed article that provides a detailed account of our research on question framing is forthcoming in the Journal of Experimental Psychology: Applied. That article can be accessed at https://doi.org/10.1037/xap0000354


35°
HOW SAFE ARE THE CONDITIONS?
WHAT EVIDENCE OF SAFETY DO I SEE?
THERE IS NO DEFINITIVE EVIDENCE OF SAFETY, SO I DO NOT BELIEVE THAT IT IS SAFE.
I DO NOT BELIEVE THAT IT IS SAFE, SO I WOULD NOT SKI IT.
HOW DANGEROUS ARE THE CONDITIONS?
WHAT EVIDENCE OF DANGER DO I SEE?
THERE IS NO DEFINITIVE EVIDENCE OF DANGER, SO I DO NOT BELIEVE THAT IT IS DANGEROUS.
I DO NOT BELIEVE THAT IT IS DANGEROUS, SO I WOULD SKI IT.
3