COLLECTIVE INTELLIGENCE IN SAFETY JUDGEMENTS BY GROUPS OF SKIERS

Tim Dassler[1,4,*], Richard Fjellaksel[2,4], Michael Morreau[3,4]

[1]*Department of Psychology, UiT The Arctic University of Norway, Tromsø*
[2]*Department of Health and Care Sciences, UiT The Arctic University of Norway, Tromsø*
[3]*Department of Philosophy, UiT The Arctic University of Norway, Tromsø*
[4]*CARE Center for Avalanche Research and Education, UiT The Arctic University of Norway*

ABSTRACT: Underestimating avalanche factors can give backcountry skiers a false sense of safety. In theory, aggregating the estimates of group members can enable a group to make assessments that are more accurate than estimates by individual members. This study finds that this effect can be of practical importance to groups of backcountry skiers. Using estimates of slope steepness by students and instructors in avalanche courses in Norway, we studied the relationship between slope steepness, group size, and the probability that the group wrongly judges a slope to be under the critical threshold of 30°. We represent the relationship between group size and the probability of a false safety judgement using safety-judgement performance curves. Such curves could be used in tour planning to answer questions such as: given the terrain, how many skiers should there be in the group in order to keep the chance of false safety judgements below a given threshold of acceptability? Or, given the group, on what kind of terrain are the group's safety judgements likely to be acceptably accurate? Improvement in safety-judgement performance as represented by the performance curves could be used also for assessing training in avalanche courses.

Keywords: Wisdom of crowds, risk assessment, avalanche education, avalanche course evaluation, trip planning, hazard management.

1. INTRODUCTION

One dangerous mistake that backcounty skiers make is to think that they are safe from being avalanched when really they are not. The chance of that depends in part on physical features of the decision environment: slope steepness and so on. When conditions are in fact extremely safe, as in fig. 1, the chance of anyone having a *false* sense of safety is next to zero: people will think they are safe from being avalanched and they will be right about that. When on the other hand conditions are extremely dangerous, as in fig. 3, the chance of wrongly thinking that you are safe is also extremely low. No one in their right mind is going to have a false sense of *safety* on this slope, because no one is going to feel safe there at all. At extremes of safety and danger, then, correctly judging safety is easy. Where it gets tricky is on slopes like those in fig. 2, where it is harder to tell whether there might be an avalanche. This is typical in the kinds of terrain that back-country skiers seek out.

Besides the physical environment, people's ability to make good safety judgements also depends on just *how* they make their judgements. One possibility for a group of skiers is to have a leader, who decides on behalf of the whole group. This is the case, for instance, in guided touring. Another is to arrive at judgements together, perhaps in discussion, or by voting, or both. Research in social psychology suggests that making judgements in a group can in itself be a risk factor. Pluralistic ignorance, where group members hide private misgivings from one another, is one of many social effects that are implicated in risky behavior Prentice and Miller (1996). Indeed, due to the challenges related to group structure and dynamics, backcountry skiing in groups of four or more often is recommended against Zweifel et al. (2016).

This article is about a positive contribution that judging together can make. Where group members feel psychologically safe, and are able and willing to share their views and decide together, groups can in theory make more-accurate judgements and reach better decisions than the individuals that make them up. This effect, known for over a century, is called collective intelligence, or the wisdom of crowds Galton (1907), Surowiecki (2004). Its theoretical relevance to avalanche safety judgements has been pointed out before Ebert and Morreau (2023). Here we confirm its relevance in practice in an empirical study of assessments of avalanche conditions by back-country skiers. Provided its benefits are not undone by pluralistic ignorance, difficulties in communicating and other detrimental social factors, we argue, making safety judgements in larger groups can be better than in smaller groups or alone.

Oversimplifying somewhat, let us say that a slope is safe from avalanches if its actual steepness is 30° or less. Similarly, we assume that people do in fact judge a slope to be safe if, in their assessment, this

*Corresponding author address:
Tim Dassler, Department of Psychology
9037 Tromsø, Norway;
tel: +47 48353465;
email: tim.dassler@uit.no

slope is 30° or less.[1] Now consider a group of skiers assessing the steepness of some slope. First, each individual member makes their own estimate of its steepness. The group reaches its collective estimate, we stipulate, simply by averaging the individual estimates of the members.[2] Finally, let there be some probability threshold of acceptability for false safety judgements. For instance, group members might be willing to accept a 5% chance of wrongly judging themselves to be safe. If the chance of a false safety judgement is greater then this, then the group decides not to ski in the terrain in question, whereas if the chance of a false sense of safety is less than their agreed threshold they are willing to go.

The questions asked and tentatively answered in this study are these: for what combinations of actual slope steepness and group size is the chance of a false safety judgement 5%? For which combinations is the chance lower than this, and for which is it higher? Finally, does training in standard slope measurement techniques appreciably improve the ability of individuals groups to judge avalanche safety?

This study has several important limitations. That's why the answers are only tentative. For one thing, it is based on limited data on slope estimates, collected during avalanche courses in just a single season. For another, our analysis of the data was done by hand, developing some of the concepts and methods as we went, and without the benefits of having a proper data scientist on the team. We think even so that the approach is useful and worth following up in a more rigorous study than we were able to make, here. We think of this article as something like "beta" in climbing: insight and information in how to take on a problem, as opposed to conclusive proof that the problem has been solved.

2. METHOD

2.1 Data collection

We collected data at both level 1 and level 2 avalanche courses in the Troms region of northern Norway between December 2022 and March 2023. At the courses students were asked to estimate ("eyeball") slope steepness from several standpoints: A, on the slope; B, from above, looking down; C, from the side and at a distance; D, looking

Figure 1: First author's two-year old daughter scoping a line. There is basically no chance of a false sense of safety, here, because conditions are in fact perfectly safe.

Figure 2: Being able to assess steepness in moderate (left panel) to complex terrain (right panel) lets you avoid dangerous avalanche trigger and run-out zones. Here someone might underestimate steepness and feel safe when really they are not.

Figure 3: Nobody is going to have a false sense of safety here simply because nobody is going to feel safe: even a small sluff avalanche will knock you off your feet and the fall will kill you. Krister Røhme Kopala in extreme terrain in Chamonix, France.

[1]This presupposes that the slope is not in an avalanche path or run-out zone.

[2]There are different ways a group can move from the estimates or judgements of its members to collective judgements of the group as a whole. For instance, instead of averaging the members´ judgements it could take the median of these. Or else it could take the lowest of them, or the highest, or the judgement of the group´s leader or something else. Center tending aggregation methods such as averaging and taking the median are known to produce wisdom of the crowds.

up at the slope at a distance; and finally E, looking up at the slope from the bottom of this slope.[3]

The instructor stopped the group and asked members to estimate the steepness of a slope. The instructor also estimated the steepness and all estimates were recorded. Then the instructor and participants measured the steepness of the slope and the measurements were recorded. For A. slopes this was done by using an inclinometer on a ski pole (either a digital inclinometer, a mobile phone inclinometer or a compass with inclinometer). For B. slopes the measurements were taken with a compass inclinometer or an inclinometer attached to a ski pole aiming down the slope. For C. slopes either a compass with inclinometer was used, or a mobile app, or a ski pole with inclinometer. For D. slopes, compass-and-digital-ski-pole-attached inclinometers were used. E. slopes were not included in the study.

Data was collected and recorded in a secure online survey by two researchers who also were instructors at the courses.

2.2 Dataset

The dataset consists of a total of 150 individual judgements (estimates together with measurements) of the steepness of 13 slopes varying in steepness from 16° to 29°. These break down into 13 instructor estimates, 13 instructor measurements, 63 participant estimates, and 61 participant measurements. Most judgements were done A. on the slope (58 percent) and C. from the side from a distance (25 percent).

3. ANALYSIS

This section discusses the concepts used to analyze the data and arrive at our results.

For the purpose of this study, we consider an *avalanche-safe slope* to be a slope whose steepness is 30° or less.[4] An *individual assessment* of slope steepness is an estimate of steepness, or a measurement of steepness, made by an individual person. A *group assessment* of steepness is the (arithmetical) mean of the individual assessments of several individuals. It is an estimate or a measurement, accordingly as all of the individual assessments it comprises are estimates or measurements. The group assessments in our study are nominal. That is, they are the averages of some number of individual assessments chosen at random from all of the individual assessments in the data set. They are not (necessarily) assessments reached by real groups of students that met together and interacted in the avalanche courses.

The *true steepness* of a slope is identified for the purpose of this study with the steepness measured by an instructor. The *error* of a an assessment, whether of an individual or a group, is the number of degrees that this assessment is below (a negative error) or above (positive) the true steepness of the slope.

For each group size from 1 to 6, and for both estimates and measurements, we produced a histogram of the numbers of errors in a total of 100 assessments, keeping negative and positive errors separate. For groups of 1 (that is, individuals) we drew at random 100 times from the set of individual estimates and 100 times from the set of individual measurements, calculated the errors, and plotted these. For a nomimal group of n members, a single estimate error was produced by sampling at random n individual estimate errors and averaging these, repeating 100 times to produce the histogram of 100 estimate errors for groups of n members. Analogously, we produced the histogram of 100 measurement errors for groups of n members.[5]

Error distributions were inferred informally from observed frequencies of errors of different kinds in the data.[6]

$\Delta_n^{5\%}$ is defined as the largest magnitude such that, for a group of n members, 5% of negative errors are of this magnitude or greater.

Suppose the steepness of a slope is in fact $30 + \Delta_n^{5\%}$°. Now a group of n skiers assesses its steepness. Assuming the error distributions for slopes just above 30° are the same as the observed distributions for the slopes below 30° used to arrive at $\Delta_n^{5\%}$, we may expect that in 5% of cases the group will wrongly assess the slope to be 30° or less.[7] That

[3] Additional data that was collected but is not used in this study includes: aspect and elevation of slopes; states of participants (tired, fresh, hungry, focused, or distracted); also visibility (high, medium, low), light conditions (sunny and clear, foggy, flat light, dark), name of mountain, date and time.

[4] For a good introduction to the role of avalanche terrain for avalanche risk and safety see Tremper (2018).

[5] The rationale for this way of doing things is that the error of an average of assessments is equal to the average of the errors of these assessments. We did not consider separately errors made on slopes of different true steepness. This lumping together of all the data gathered on different slopes was desirable given the fairly low number of individual assessments that we had. It is a weakness of the study though, since the slopes in question differed considerably in their steepness, from 16°to 29°. There is no good reason for considering only 100 judgements. It would have been better to consider thousands, but lacking the computational expertise we did most of the analysis by hand. 100 was all we had time for.

[6] That is, having calculated and plotted a histogram of the errors of individual and collective judgements, we sketched by hand our best guess as to the error distribution responsible for the observed errors. This might be regarded as acceptable in a "proof of concept" study such as this.

[7] Distributions of errors on slopes not too far above 30° should be quite similar to error distributions on slopes not too far below 30° if the 30° threshold, though significant for the physics of avalanches, has no special significance in the visual perception and psychology of humans. This assumptions seems reasonable to us though we have no experimental evidence to support it.

is, on the earlier identification of avalanche safety with a slope's being in fact 30° or less, and of safety judgements with assessing the slope to be in this sense safe, the chance of that the group wrongly judges itself to be safe is 5%. Figure 4 plots the relationship between group size n and $30 + \Delta_n^{5\%\circ}$ for a range of realistic group sizes in backcountry ski touring.

4. RESULTS

The safety-judgement performance curves of figure 4 summarize the result of our analysis.

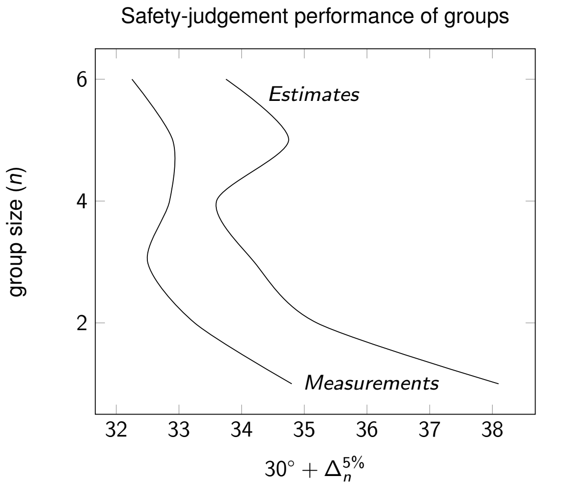

Figure 4: Combinations of group size and slope steepness for which the chance is 5% that the group wrongly judges the slope to be safe (30° or less) on the basis of individual steepness estimates and measurements. To the top right of either curve the chance of a false safety judgement is less than 5% and to the bottom left it is greater.

Here´s how to read the curves. Take a point on either of the curves, say (34.2, 3) on the *Estimates* curve. This point tells us that a group of three skiers that estimates slope steepness (rather than measuring it), assessing a slope whose true steepness is 34.2°, may be expected to have a 5% chance of wrongly estimating this slope to be less than 30°. This we count as wrongly thinking that this slope is safe. If assessing a slope whose true steepness is less than 34.2°, they are even more likely than this to assess it as below 30° and safe. Given that 5% is taken as the threshold for acceptability of this kind of mistake, the group of three should avoid terrain in the range 30° – 34.2°, where they're too likely to feel safe when actually they're not. Either that, or else they should find some better way of assessing slope steepness.
One better way, we see from figure 4, is to measure steepness rather than merely estimating it. Move over to the *Measurements* curve and find the point (32.5, 3). This point tells us that a group of three that measures slopes has a 5% chance of a false safety judgement when the true steepness is 32.5°. When assessing slopes that are steeper than this, the group has less than a 5% chance of a wrong safety judgement. In particular, this group of measurers has an acceptable (less than threshold) chance of false safety judgement when considering slopes between 32.5° and 34.2°. A group of three estimators, on the other hand, merely "eyeballing" the slope, has an unacceptable (more than threshold) chance of false safety judgement for these slopes.
Measuring, as taught in the avalanche courses in which we collected data, apparently is a better way of assessing steepness than "eyeballing". It demonstrably increases the range of terrain in which backcountry skiers can acceptably operate. This illustrates our proposal that safety-judgement performance be used to evaluate th effectiveness of avalanche courses in teaching techniques such as slope measurement.
Another way to improve judgement is to increase the size of the group. On the *Estimates* curve we find the point (33.6, 4). Considering a slope above 33.6°, a group of four may be expected to have less than a 5% chance of a false safety judgement. In particular, it has an acceptable chance of false safety judgement when considering slopes between 33.6° and 34.2°. So increasing group size by one will improve judgement and increase the range of terrain in which the group can acceptably operate — though not as much as measuring would.
We emphasize that these results are tentative, given previously mentioned limitations of the present study.

5. DISCUSSION

5.1 Wait, aren't group decisions FUBARed?

This study shows that groups of three or more are better at estimating slope steepness than individual skiers. They are thus less likely to make false safety judgements and wrongly judge a slope to be 30° or less when it is not. This is due to a wisdom of crowds effect Galton (1907); Surowiecki (2004).
The idea that it is bad, even dangerous to make decisions in groups comes from thinking about other things than wisdom of crowds that affect decision quality. The quality of leadership and communication among group members, as well as the kind of decision-making process Bright (2010); Cierco and Debouck (2013), are important here. So are heuristic traps, risk tolerances, gender charcteristics and many other matters impacting what people do and whether they choose potentially hazardous terrain Hendrikx and Johnson (2016); Hendrikx et al. (2022).
Whether it is in actual practice better for backcountry skiers to decide in groups depends on the balance of the effects of the wisdom of crowds and these other factors. How the different factors go together in groups of skiers has not been studied here.

5.2 What's up with the larger groups?

Figure 4 shows improvement with increasing group size up to a certain number of people, 3 and 4 for measurements and estimates respectively, but with more people the curves turn back. We do not know why this happens. Perhaps it is just that our data set is too small and what we´re seeing here is simply noise. In this case, the puzzling reversal strengthens our belief that the method proposed here should be used to conduct a larger and more powerful study, perhaps of other avalanche factors than steepeness such as snowpack properties and weather conditions. Think of this study as a pilot, showing the way.

5.3 How can there be less risk on steeper slopes?

We show that the chance of making false safety assessments is lower when moving on steeper slopes (say, 38°) than on less steep ones (say, 33°). This might seem a bit of a paradox, since 38° is right around the sweet-spot for triggering avalanches. How can steeper be less risky? It is important not to misunderstand us as saying that you are safe when you go 38° or steeper. What we show that on the steeper slopes you´re in a better position to *know* whether you are in danger or not. Having this knowledge can reduce avalanche risk, understood to factor in not only the probability of avalanches but also exposure to harm: the better one is at knowing whether or not one is in danger, the better one will be at keeping out of harm´s way.

5.4 Just how smart are small crowds of skiers?

Groups of 3-5 skiers that guess steepness seem to be about as accurate as single skiers measuring it. Figure 5 compares the error distributions of four people guessing (red graph) with one person measuring (blue). The two distributions are similar. Evidently, aggregation has boosted the accuracy of collective guessing to the point that it is about as good as individual measuring. The wisdom of crowds promises to be quite useful with groups of sizes that are realistic in backcountry skiing.

6. SHORTCOMINGS AND LIMITATIONS

Due to ethical and safety reasons data on estimates was gathered on slopes below 30 °and extrapolated to steeper slopes. The rationale is that whereas 30 °is critical for avalanche risk, it isn't for human perception of slope steepness. This might be reasonable for slopes just about 30 °but we don't know how reasonable it is for steeper slopes.
We used informal methods to estimate the distribution of assessment errors from observed frequencies of errors. This might be regarded as suitable in a pilot study such as this, but a more rigorous and repeatable process is required for a robust study.

Figure 5: The number and magnitude of errors for one person measuring (blue graph) and a group of four people guessing (red graph), with 100 judgements of each kind. The number and size of the errors are in both cases quite similar.

7. CONCLUSION AND IMPLICATIONS

Current best practice is to tour in smaller groups. Our preliminary finding is that larger groups need not always be worse when it comes to making more precise safety judgements. Because of a wisdom of crowds effect, four touring buddies guessing the steepness of a slope may show better performance than a solo tourer measuring. This has implications for both backcountry recreationists as well as avalanche education.
If you are out in potentially hazardous winter backcountry terrain it can be beneficial to bring touring buddies. This will increase performance of judging slopes to be safe. Note that this presupposes that the group actually is able to communicate their judgements to each other. In group settings where it is socially awkward or difficult, or with individuals who lack the communicative skills and courage to speak up, this effect might be canceled out.
The study may also be used to create evaluation methods for avalanche courses to check whether or not course participants got better at estimating and or measuring slope steepness. If steepness judgements are collected in the beginning and at the end of a course, and analysis is automated, performance envelopes may be used to show learning, that is, improvement in assessing steepness, for both individuals or groups.

ACKNOWLEDGEMENT

We thank for their comments, suggestions and other help Frank Techel, Thorleif Svendsen, Espen Minde and Markus Johnsen Aase, as well as members of the GUESSED project (Nordforsk project number 105061). Thanks also to Krister Røhme Kopala for the picture of him in some pretty hair-raising terrain.

References

Bright, L. S.: Group dynamics and decision making: Backcountry recreationists in avalanche terrain, Colorado State University, 2010.

Cierco, F.-X. and Debouck, F.: Human factors in decision making in avalanche terrain, in: Proceedings of the International Snow Science Workshop, Grenoble, pp. 7–11, 2013.

Ebert, P. A. and Morreau, M.: Safety in numbers: how social choice theory can inform avalanche risk management, Journal of Adventure Education and Outdoor Learning, 23, 340–356, doi:10.1080/14729679.2021.2012216, URL https://doi.org/10.1080/14729679.2021.2012216, 2023.

Galton, F.: Vox Populi, Nature, 75, 450–451, 1907.

Hendrikx, J. and Johnson, J.: Understanding global crowd sourcing data to examine travel behaviour in avalanche terrain, in: Proceedings of the 2016 International Snow Science Workshop, pp. 737–743, 2016.

Hendrikx, J., Johnson, J., and Mannberg, A.: Tracking decision-making of backcountry users using GPS tracks and participant surveys, Applied Geography, 144, 102 729, 2022.

Prentice, D. A. and Miller, D. T.: Pluralistic Ignorance and the Perpetuation of Social Norms by Unwitting Actors, vol. 28 of *Advances in Experimental Social Psychology*, pp. 161–209, Academic Press, doi: https://doi.org/10.1016/S0065-2601(08)60238-5, URL https://www.sciencedirect.com/science/article/pii/S0065260108602385, 1996.

Surowiecki, J.: The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations, Little, Brown, 2004.

Tremper, B.: Staying alive in avalanche terrain, Mountaineers books, 2018.

Zweifel, B., Procter, E., Techel, F., Strapazzon, G., and Boutellier, R.: Risk of avalanche involvement in winter backcountry recreation: the advantage of small groups, Wilderness & environmental medicine, 27, 203–210, 2016.