Journal of Adventure Education and Outdoor Learning

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/raol20

Safety in numbers: how social choice theory can inform avalanche risk management

Philip A. Ebert & Michael Morreau

To cite this article: Philip A. Ebert & Michael Morreau (2022): Safety in numbers: how social choice theory can inform avalanche risk management, Journal of Adventure Education and Outdoor Learning, DOI: 10.1080/14729679.2021.2012216

To link to this article: https://doi.org/10.1080/14729679.2021.2012216

JOURNAL OF ADVENTURE EDUCATION AND OUTDOOR LEARNING
https://doi.org/10.1080/14729679.2021.2012216

Safety in numbers: how social choice theory can inform avalanche risk management

Philip A. Ebert[a,b] and Michael Morreau[c]

[a]Division of Law and Philosophy, University of Stirling, Stirling, UK; [b]Behavioural Science Centre, Department of Economics, University of Stirling, Stirling, UK; [c]Philosophy Department, University of Tromsø, Norway

ABSTRACT

Avalanche studies have undergone a transition in recent years. Early research focused mainly on environmental factors. More recently, attention has turned to human factors in decision making, such as behavioural and cognitive biases. This article adds a *social* component to this human turn in avalanche studies. It identifies lessons for decision making by groups of skiers from the perspective of social choice theory, a sub-field of economics, decision theory, philosophy and political science that investigates voting methods and other forms of collective decision making. In the first part, we outline the phenomenon of *wisdom of crowds*, where groups make better decisions than their individual members. Drawing on the conceptual apparatus of social choice theory and using idealised scenarios, we identify conditions under which wisdom of crowds arises and also explain how and when deciding together can instead result in worse decisions than may be expected from individual group members. In the second part, we use this theoretical understanding to offer practical suggestions for decision making in avalanche terrain. Finally, we make several suggestions for risk management in other outdoor and adventure sports and for outdoor sports education.

KEYWORDS

Risk management; outdoor sports; group decision making; social choice theory; Condorcet's jury theorem

1. Introduction: knowledge and ignorance in groups

Avalanche studies have undergone a transition in recent years. Early research focused mainly on environmental factors such as the physical aspects of different snow profiles, avalanche dynamics depending on terrain and snow-type, and the influence of weather-conditions on snow stability. More recently, attention has turned to human factors in decision making such as behavioural and cognitive biases. This article adds a *social* component to this human turn in avalanche studies. It identifies lessons for decision making by groups of skiers from the perspective of social choice theory, a sub-field of economics, decision theory, philosophy and political science that investigates voting methods and other forms of collective decision making. The focus throughout is on group decisions in avalanche terrain. The main insights, however, are quite general and relevant to risk management in other outdoor sports as well.

There are obvious advantages to being in a group when venturing into dangerous terrain. For one, when someone gets into trouble there are others around to help. There can also be real disadvantages, though. Adverse effects of group dynamics on decision making have been identified as a contributing factor in several accidents, including a 2007 accident on the Jungfrau in Switzerland (Harvey, Winkler, Techel, & Marty, 2013), which resulted in six fatalities, and the widely publicised 2012 Tunnel Creek accident in Washington State, U.S.A., with three fatalities (Branch,

CONTACT Philip A. Ebert p.a.ebert@stir.ac.uk Division of Law and Philosophy, University of Stirling, Stirling, FK9 4LA, UK

2012). Some of the mechanisms by which group dynamics can make for bad decisions are by now well known. McCammon in a famous study identified different heuristic traps for individual and group decision making (McCammon, 2004, 2009). One of these is the *expert halo*, where someone is perceived to be an expert on the basis of characteristics that are not necessarily good indicators of expertise, such as age or even a good track record of avoiding avalanches (Ebert, 2015; Ebert & Photopoulou, 2013). Another pitfall is *group polarization*, which can lead to an increase in risk-seeking behaviour (Sunstein, 2000). Several experimental studies have (broadly speaking) supported McCammon's findings (Furman, Shooter, & Schumann, 2010; Leiter, 2011; Marengo, Monaci, & Miceli, 2017). In addition, a study by Zweifel and Haegeli (2014) has offered important empirical insights into how, specifically in the context of backcountry skiing, group formation and leadership tend to develop. These insights identified 'feeding' grounds for the development of specific biases in groups.

The first aim of this article is to explain how deciding in groups can contribute to the safety of groups. Under some conditions it can be expected to amplify the knowledge and expertise of individuals, enabling groups to make better decisions than could be expected from any of the members on their own. This *wisdom of crowds*, or *collective intelligence*, is widely discussed in economics and political science, having practical applications in medical decision making, management, and predictions models (for a popular introduction, see (Surowiecki, 2004)). Its importance for risk management in outdoor sports decision making has until now not been widely appreciated.

However, when the conditions for collective wisdom are not met, deciding in groups can amplify the ignorance or rather incompetence of group members, resulting in collective decisions that are likely to be worse than those of individual members. Deciding in groups, under such circumstances, is itself a *risk factor*. A second aim of this article is to identify and explain a critical notion of individual decision-competence, and to show why it can be hard to know whether people are competent in the relevant sense–even if they are highly experienced decision makers.

Finally, a third aim of this article is to offer implementable suggestions for current best practise. We separate different stages of the group decision-making process and show how our discussion can inform outdoor sports risk management.

There are two basic scenarios in which collective intelligence arises. The first is with majority voting. Suppose there's a choice between two options, and that one of these is in some relevant way the *right* one to choose. Some people are going to make the choice, and they're somewhat uncertain about which option this is. A famous result from political science, the *Condorcet Jury theorem*, tells us that, under certain conditions, majority voting is a way of increasing the chance of choosing the *right* option (Condorcet, 1785; List, 2013). The second scenario for collective intelligence is when people estimate the value of some quantity such as risk, that comes in degrees. In a paradigmatic case, Sir Francis Galton, one of the founders of statistics, showed by analyzing a competition to judge the weight of a prize ox at a county fair that the median or middlemost judgment was closer to the true weight than, on average, the judgments of individual group members (Galton, 1907). More generally, this scenario is relevant also for decisions made using scores or grades, such as for instance numerical risk scores, and a jury theorem is available then as well (Morreau, 2021).

The structure of the article is as follows: in the next section, we introduce idealised scenarios ('thought experiments') in which collective intelligence could improve a decision by a group of skiers in avalanche terrain, and we explain why individual *decision-competence* is an important precondition of wisdom of crowd effects. Here, we also discuss when it is appropriate to use majority voting and when a grading approach is more suitable. In the subsequent section, we discuss another requirement for collective intelligence, *independence*, and show why and how it can be met in an outdoor sports context. Here, we also discuss how *deliberation* can help individuals make better decisions, and when it can become a risk factor. For the sake of readability, we relegate to appendices most of the technical results from social choice theory that underlie our discussion. In sections 4 and 5, we then identify practical implications based on our understanding of social choice theory for current best practice and show how our discussion is relevant to outdoor education and risk management in outdoor and adventure sports more generally.

2. Group wisdom: individual and group decision-competence

Say a group of skiers is out on a trip and there are two options to choose between: they can ski down the northern slope, let us say, or else down the southern one. One of these options is in some relevant way the *right* one to choose. Supposing for the sake of the example that they care only about their safety, that would be whichever slope is the safest one.[1] Now, suppose everyone in the group is a well-educated and experienced decision maker with good knowledge about avalanche safety. Assume further that each member has diligently collected specific information about the snow stability of both slopes and that, as a result, each one individually has a pretty good chance of reaching the right judgment, let's say it is 80%. This means that any given member has an eighty percent chance of making the right choice on their own, in this case. In the long run, in cases just like this one, you'd expect the right decision eight times out of ten. This probability of getting it right on a specific occasion is what we call the individual *decision-competence*.

Now, the different group members may disagree about what to do, and so they'll need some way to resolve disagreements. One way for the group to arrive at a collective decision, also in cases where there is disagreement, is for them to have a leader who decides on everyone's behalf. Another way is to have a majority vote: if most members think the slope is safe, then that's the group's decision. Importantly, majority voting can amplify the decision-competence of group members. With three skiers, each with an eighty percent chance of getting the decision right on their own, the chance that the judgment of the majority is right is, under favourable conditions, 90%–a ten percent improvement on the decision-competence of individual group members. The larger the group the better, and with five members the collective decision-competence under majority decision can be as high as 94% (see appendix A).[2]

One crucial requirement for collective intelligence to emerge through majority voting is that the decision-competence of individual group members is above 50%. Where it is not, and individual voters are more likely to be wrong than right, the knowledge machinery shifts, so to speak, into reverse gear. Then the majority is even more likely to make the wrong decision than the individual members, and the more members there are the greater is the chance of the majority settling on the wrong option. This is a further mechanism by which collective decision making can be a risk factor, in addition to the expert halos, group polarization, and biases mentioned earlier.

Among the factors that tend to decrease individual decision-competence are ignorance and carelessness in thinking through all the different things that could go wrong on a trip (compare the findings by Zweifel and Haegeli (2014)). But decision-competence can slip below the 50% mark even when people do their best to gather all the relevant information, and to take everything into account. Suppose, for instance, that a so-called weak layer on which an avalanche could be triggered, identified by the group in their planning, is in fact far more reactive than any testing during the trip has suggested. In this case, the group's evidence about the stability of the slope is badly misleading. As a result, the decision-competence of the members of the group–their chance of giving the right answer *in this particular case*–is not 80%, but lower than that. Let's say that under these circumstances it is only 20%, which means that they are more likely than not to be wrong: in the long run, under relevantly similar circumstances, each one makes the right call just two times in ten. Now, majority voting about whether the slope is safe makes things worse. The probability that the majority gets it right is less than the probability that an individual member of the group does.

More generally, the point is that majority voting amplifies individuals' *knowledge* only if they know enough to begin with. If they don't know enough, then majority voting amplifies their *ignorance* instead. Importantly, because people's individual decision-competence can vary around the critical 50% mark depending on factors that are unknown to them (here, the unknown reactiveness of the weak layer), people in general cannot be certain that they meet the decision-competence threshold so that majority voting will result in better decisions. They might *think* they are competent to decide (because they are such experienced skiers, have often skied this slope without any problems, and so on) but that is another thing. What's at issue is not their *general* competence and skills to decide in a range of different cases, but their individual *decision-competence*, that is, their probability of getting the right answer under the particular conditions in which they find themselves that day, which might or might not be in important respects unknown to them. So then, what are the conditions under which it is reasonable for groups to use majority voting as a risk management tool?

One upshot is that it's not enough for the skiers to be thinking about what's out there in the world, that is, the condition of the snow. In addition, they have to be thinking about the state of their own knowledge–or, as the case may be, their ignorance–about what's out there in the world. There might be 'known unknowns': for instance, on reflection they might realize that the evidence they collected isn't enough to know how reactive the weak layer in question is, and thus realize that they do not know whether they have the decision-competence needed to make a good decision. When individual members of the group are reasonably confident that they are competent in the specific scenario because they know there are no unknowns, for example, because they know the snowpack is consistent with no spatial variability, or because they know they are confronted with an easily testable avalanche problem, and so on, then majority voting can help them to make better decisions. When, on the other hand, they are not sufficiently competent in the specific scenario (which might come as an unwelcome surprise to them, just like the reactiveness of the weak layer) deciding together in this way can be dangerous. Knowing that you are competent is thus an important factor in responsible decision making not just for individuals but also for groups. Soon we will consider other ways of deciding in a group that might be less sensitive to variations in decision-competence.

First, though, let's look at a way in which aggregating different inputs can help to increase decision-competence by means of another example. Consider the case in which the skiers are going to look into some of the 'known unknowns', for example, by performing a snow stability test. They opt for a so-called Extended Column Test (ECT).[3] It requires little interpretation and, using some simplifying steps, results in a binary output: stable or not-stable. The reliability of an ECT for either output is, we may assume, over 50%.[4] Now the group can treat each of their ECT results as an individual vote, and take the aggregated result of several ECTs to be the result that is gotten in most of them. Just as with people, the chance that the outcome reached in a majority of ECTs is right is greater than the chance that any single ECT gives the right result.

There is a complication: supposing they do four tests with two suggesting a not-stable slope, and two suggesting a stable one? One obvious solution is to have a tie breaker. If one location is more relevant than another, more weight might be placed on that stability test. Or if one person is known to be more skilled in performing the test, this could tip the balance. Failing this, another test might be required. The same complication arises when it is people who are voting, and there is no majority for any of the options. Here again, a tie breaker could be the most experienced or skilled person. From an evidential perspective, tipping the balance in this way would be the best way forward. However, from a group management perspective in a high-risk situation this could be the wrong thing to do. If two out of four people are judging a slope as not-safe that may be enough to decide to play it safe. We come back to this issue of how evidential constraints and other ethical or practical constraints can pull in different directions in section 4.

So far, we've considered a decision in which there are just two options: safe and not-safe, go or not-go. In some practical cases there are more than two options to consider. Suppose our group of skiers has arrived at the top of the mountain, and imagine they have three descent options: They can ski slope A, slope B, or slope C. How, ideally, should the group come to a joint decision in such a case? One possibility is to break the choice among several options down into several binary choices, by taking the options in pairs. Supposing most group members were to judge that slope A is safer than slope B, for instance, and a majority were to judge that slope B to be safer than slope C, then we would take it to be the judgment of the group as a whole that slope A is the safest of the three.

There is a problem with breaking multiple-option decisions down into binary decisions. It can lead to incoherent results. To see how, suppose there are three skiers in the group, and that their judgments of the relative safety of the slopes happen to be as follows: Anna thinks that slope A is safest, followed by slope B and then slope C. Bertil thinks that slope B is safest, followed by slope C and then slope A, while Carsten thinks that slope C is safest, followed by slope A and then slope B. Both Anna and Carsten vote for slope A above slope B, so the group opinion is that:

(1) Slope A is safer than slope B.

Also, a majority (Anna and Bertil) thinks that slope B is safer than slope C, so the group thinks that:

(2) Slope B is safer than slope C.

If the group had a coherent judgment about safety then it would also think that:

(3) Slope A is safer than slope C,

since (3) follows immediately with simple reasoning from (1) and (2). In fact though the group's opinion is just the other way around, since both Bertil and Carsten think that slope C is safer than slope A. Though individually the group members' judgments about the relative safety of the slopes are perfectly coherent, the collective judgments reached by majority voting about pairs of options are not coherent. This way of approaching a decision with several options can turn rational individuals into irrational groups. This result is sometimes known as the *Paradox of Voting*, or *Condorcet's Paradox* (List, 2013).

Fortunately, there is another way to make multi-option decisions that is guaranteed to produce coherent outcomes. It is to score or grade the options, first individually and then collectively, and to make decisions on the basis of collectively assigned grades. In the earlier example, suppose that instead of simply comparing the different slopes with respect to how risky they are, each of the skiers evaluates each of the slopes separately by assigning it a score, or a grade. For instance, these could be Likert scores on a scale $1 - 5$, where 5 is the highest level of risk ('extremely dangerous'), 4 the next highest level, and so on down to 1, the lowest level ('safe as houses'). Now the three of them will decide together by evaluating each slope separately, and comparing the results. One natural way to do this is to let the collective judgment concerning any particular slope be the middlemost or *median* judgment–the one in the middle when everyone's scores are put in order.[5] If Anna gives slope B a 4, Bertil a 3 and Carsten a 5 then the group assigns the risk score 4 to that slope. If reasoning in the same way they arrive at the collective score 3 for slope A then they collectively judge slope B to be the riskier slope of the two, since its risk score is higher. This decision method always produces a coherent result.[6]

Scoring and grading the options instead of holding majority votes among them can make a further important contribution to the collective intelligence of groups of skiers. We have seen that majority voting entails an increased risk of making the wrong decision when the individual chance of making the right decision drops below the critical level of 50%. That can happen when voters are subject to misleading evidence but also when they are biased. Now when such decisions are made differently, by scoring each of the options separately, first individually and then collectively, right decisions are in a range of cases forthcoming even so. Groups of skiers that decide by scoring their options are sometimes less susceptible to be led astray by individual biases or misleading evidence than are groups that decide by majority voting. A case of this sort is sketched in appendix C.

We have outlined two decision-making procedures for groups: majority voting in binary choice, and median voting when grading or scoring different options. We noted some of the pitfalls for group decision making, such as a failure to meet the decision-competence condition. Next, we look at the second requirement, independence, for harnessing the wisdom of a group of skiers.

3. Group dynamics: the interplay of independence, deliberation, and decision-competence

Decision-competence is one factor that underwrites the wisdom of crowds. Another is the absence of certain dependencies among the decisions of group members. More specifically, the Condorcet Jury theorem not only requires that individual decision-competence is above 50%, but also that individual decisions are *statistically independent*. This means, in technical terms, that the likelihood of any given group member's deciding this way or that is the same, no matter which decision is made by any other member of the group. The technical characterisation is, however, easily misunderstood. In particular, it is compatible with the independence assumption that individuals are coordinated in some way. Competent voters are, after all, coordinated (to some extent) by the facts or by a shared body of evidence. While facts or evidence may influence and co-ordinate, individual decisions may still count as independent provided that the way in which decision makers can make errors is not coordinated, that is, individuals may make mistakes for different reasons.

To summarise, for collective intelligence to arise we need both decision-competence and independence. Decision-competence can be understood as a relation between an individual and the world: the individual is sufficiently good in predicting the world. Independence can be understood as a relation amongst the individuals of a group: each individual is resilient in their opinion given the other group members' opinions.[7] An immediate consequence of this notion of independence is that if a group of skiers *blindly and thoughtlessly* follow the decision of the perceived expert, then their judgements are not independent and there is no possible improvement nor any possible worsening of the group's decision making over that of the perceived expert. If, however, a group of competent decision makers are all influenced by the same body of evidence yet some members of the group are prone to some biases, errors or misinterpretations that others are not and vice versa, then the judgements can still be in the relevant sense independent and a majority vote can cancel out the noise created by these biases.

Now, realistically, skiers will deliberate within a group about the evidence before they make a decision. How then should we understand the interplay between group deliberations, individual decision-competence, and independence? More specifically, does deliberation undermine independence? To highlight some of the potential pitfalls involving deliberation, let's consider, another group decision scenario with uncertainty and high stakes that has been studied in more detail, namely, jurors in a criminal trial. The analogy is as follows: while jurors aim to reduce the *risk of a false conviction*, skiers aim to reduce the *risk of a false 'safe slope' decision*, i.e. they aim to avoid judging a not-safe slope as safe. Of course, there are differences between the two groups and we come back to these later; but drawing this analogy is informative in the following ways.

First, both jurors and skiers make their judgements based on a shared body of evidence. In the good case, the evidence is reliably connected to the facts; in the bad case, the evidence will be misleading. However, only because individuals share a body of evidence, doesn't mean that other factors can't affect individuals differently, or that individuals can't assess the evidence in different ways. Hence there is, for both groups, a potential for a wisdom of crowds effect given sufficient decision-competence.[8]

Second, jurors like skiers usually deliberate amongst themselves before reaching a group decision. The notion of deliberation we are using here is meant to include more than simply openly sharing evidence. If one skier shares their knowledge of a collapsed snowpack or other bits of evidence, this does not in itself constitute a genuine deliberation. Rather, we mean with deliberation the process that once the evidence is shared the group weighs up openly the given body of evidence. They deliberate the relevance, the strength, and the weaknesses of the gathered evidence. In the good case, deliberation amongst jurors or skiers can help to reduce or eliminate errors and thereby increase individual decision-competence. In this way, deliberation can, so to speak, add extra power to the knowledge machinery of group intelligence. But again, that extra power renders collective decision making all the more dangerous when the machinery shifts into reverse: biases introduced through deliberation can pull individual decision-competence below the critical 50% threshold.

Third, just as we need to be confident–know that we know–that the decision-competence condition is met, we need to be confident that deliberation will work in the desired way for a given group. Research in group and, more specifically, jury decision making can offer us some pointers for the conditions under which deliberation can become a risk factor. Consider the phenomenon called *informational cascading*: if individual decisions are made *sequentially and openly* during an open deliberation phase then early judgements in the sequence may directly influence the decisions of individuals later in the sequence (Anderson & Holt, 1997; Bikhchandandi, Hirshleifer, & Welch, 1992). This, of course, need not in itself be irrational. After all, that someone else in a group made a certain judgment might well count as new evidence for this judgment, but it can exert too much influence and lead individuals to ignore their own private assessment. For example, in groups where decisions reflect on ones own reputation, individuals may be inclined to weigh conformity with other group members' views more strongly than they should.[9] Couple these considerations with what is known about gender in jury decision making, for example, that male members tend to speak more than female members, and that male as well as higher social status members of a jury engage at higher rates in discussion (Hastie, Penrod, & Pennington, 1983; Marder, 1987), and one can see how some unbalanced forms of deliberation can negatively affect individual decision-competence.

There is, as far as we know, little research that investigates the effects of deliberation in the case of outdoor group decision making (Zweifel and Haegeli 2014). Also, there may be important differences between juries and some outdoor groups that should make us wary of extrapolating insights from the former to the latter. For example, many outdoor groups involve members that know each other well, where decisions are made on a background of shared values, trust, and mutual respect without concern for ones reputation (or so some may think). Hence, the specific set-up of a group, that is, how it is formed, who is involved, and how diverse it is, really matters; and it is difficult to offer generic advice to improve the effects of deliberation. Nonetheless, one easily implementable strategy to limit the potential negative effects of information cascading, perceived expert bias, or the reputation bias, is to introduce a decision strategy for groups that ensures that the deliberation phase and the decision phase are clearly separated and that decisions are made *non-sequentially and anonymously*. The next part of this article will synthesize these insights and bring them to bear on current best practice in avalanche group decision making.

4. Learning from social choice theory: best practice for group decisions in avalanche terrain

Lessons from social choice theory about how, *ideally*, individuals should come to a group decision, can offer a different perspective on existing best practice and provide new ideas and suggestions to improve group decision making in avalanche terrain. Now, while our suggestions are theoretically grounded in social choice theory and informed by our discussion about the relevance of decision-competence, independence, and the effects of deliberation, they are not empirically validated in the specific context of avalanche group decision making. This much should be acknowledged as a limitation of our approach. We hope that our proposals will inspire some empirical studies.[10] Let's again consider our group of skiers who venture out on a trip and discuss separately the different stages in forming a group decision.[11]

Stage 0: gathering the evidence

Current best practice, see, for example, Tremper (2018), encourages all individuals in a group to *share the evidence* they identify and point out any possible 'red flags', such as signs of recent avalanches, new snow, collapsing snow pack, shooting cracks, temperature fluctuation, drifting snow, etc.

Lessons from social choice theory provide further theoretical grounding for this practice. According to our discussion, it is important that all relevant evidence is widely shared amongst group members, so as to enable members to make informed individual decisions which can lead to an even better group decision. Having said this, given potential pitfalls in deliberation, members of a group should be discouraged from discussing in detail the relevance or weight of the evidence before the decision problem is clearly identified.

Stage 1: framing the decision problem

Current best practice encourages groups to agree, as part of their planning, on so-called 'key places'. These are specific locations where decisions ought to be made before proceeding.[12] Our discussion suggests that the *framing* of the decision problem should be part of a 'key places' discussion. Informed group decisions can only be made when all individuals agree on what the specific decision problem actually is. As discussed in section 2, groups need to decide whether the problem involves a simple binary choice (safe vs not-safe; go vs not-go), or whether there are multiple options that require a scoring or grading approach.

Another lesson from social choice theory is that group members have to be sure to answer the *same* decision problem. Otherwise, there can't be a wisdom of the crowd effect. So, for example, there is an important and often implicit ambiguity when making risk or safety judgements: we should distinguish between a *factual risk* judgment and a *value risk* judgment. A factual risk judgment concerns, for example, the question how *likely* it is that a slope will avalanche. In contrast, a value risk judgment is one in which the risk is understood as relative to a benefit, that is, whether the relevant factual risk is worth taking. Sometimes people use phrases such as 'a slope is "sufficiently" safe', indicating that a residual factual risk is worth taking or acceptable. So, even when individuals agree that it is highly likely that a given slope is stable, they can still legitimately disagree about whether the agreed upon factual risk is actually worth taking, that is, they may still disagree about the value risk.[13]

We thus recommend when making key decisions that the decision problem is clear. If in doubt, groups can choose to make two decisions, one involving a factual risk, the other a value risk frame. Ultimately, for groups with a broadly shared set of values and risk attitudes, there will rarely be differences between the two judgements; for groups whose members have different sets of values or distinct attitudes to risk, variations may easily arise.

Stage 2: check your competence

An important best practice in the aviation industry is the *IM SAFE* checklist that pilots have to pass before they are considered fit to fly.[14] The checklist ensures that pilots meet the *minimal* conditions for being fit decision makers on that day. A similar 'competence check' could be used before individual skiers make a group choice: individuals need to reflect on whether they are sufficiently confident that they meet the decision-competence conditions as outlined in section 2.

As our discussion has indicated, the intended notion of decision-competence is case-specific: being a well-educated decision maker with a good track record is not by itself enough to count as a competent decision maker in a specific case.[15] What is required for individuals to be decision-competent is for them to be more likely than not to make the correct decision *given the specific decision problem and given the specific body of evidence*. Of course, assessing whether one is sufficiently competent in a specific case can be very difficult.[16] Earlier we suggested that individuals should focus not merely on what they think they know in a given case but also reflect on the known unknowns. What are such known unknowns?

For example, when considering slope stability, known unknowns may be remaining uncertainties about the validity of specific stability tests. This will involve considering whether the chosen test location is representative, whether the expected spatial variability of the weak layers reduces the test's relevance, or indeed whether one has actually performed the test appropriately in this case. Another set of known unknowns is the possibility of misjudging the evidence or being misled by the evidence. So, for example, skiers will sometimes reason from an *absence of evidence* of instability to *evidence of an absence* of instability, i.e. evidence of stability. This is a questionable inference that isn't, however, always inappropriate. It will depend on the specifics of the given case, in particular, on whether sufficiently many tests were performed to create a wide variety of opportunities for instabilities to appear. Also, if a judgment about the stability of a slope is based on the proven stability of a relevantly similar slope, known unknowns may be identified by considering potential *dissimilarities* between these slopes that can make a difference to their stability (such as exposure to local wind effect, sun radiation, etc.).

Lastly, reflection on how well the acquired body of evidence coheres with the predictions made in planning and in the avalanche forecast can help one realise that there are still too many unknowns. An assessment should not merely be based on the latest and most recent information, but include prior information based on avalanche forecasts and observations made throughout the day. A complete body of evidence should create a *coherent narrative* about the snow conditions. A lack of coherence suggests that there are still some unknowns. In short, understanding and questioning intricate relations between evidence and ones final assessment should be part of a competence check.

In the end, the lesson for current best practice from a social choice perspective is that the more complex the avalanche situation, the more difficult it is to evaluate the known unknowns and the more challenging it is to be justifiably confident in ones individual decision-competence. Importantly, note that the complexity of the situation given the evidence is not necessarily proportional to the underlying avalanche hazard rating. When the avalanche risk is high, decisions may be quite easy. However, when the risk is moderate or considerable with, for example, a highly variable snow-pack, evidence of instability is more difficult to detect and is more often misleading. In the end, doubts about ones own decision-competence suggest doubts about whether an informed and responsible decision can be made: the more known unknowns, the greater the gamble, and the more dangerous group decisions can become.

Stage 3: initial decision

Best practice in avalanche decision making requires considering whether any heuristic traps may inadvertently affect ones decisions, that is, groups are encouraged to go through the widely discussed FACETS checklist.[17] Drawing on social choice theory, we can offer a theoretical explanation of some of these potential traps and offer insights in how to avoid them.

The (E)xpert halo is, from a social choice perspective, a potential trap insofar it undermines independence. So, instead of harnessing the wisdom of crowds, the quality of the group decision depends only on the decision-competence of one individual. As our discussion showed, this isn't necessarily a bad thing; what is bad, however, is that an expert halo will deprive a group of the potential to do better. Furthermore, it can also lead to unwanted side effects: if there is a perceived expert who is responsible for the group decision, other individuals may well be not as attentive to identifying and sharing all the relevant evidence. So, the expert halo, in effect, reduces group decision to one decision maker with possibly less evidence.

Similar lessons from our discussion apply to other heuristic traps. (F)amiliarity with the terrain may influence individuals to ignore relevant evidence or be too heavily influenced by their previous experience, and (T)racks may provide misleading evidence about slope stability. In short, both 'heuristics' can reduce individual decision-competence; however, and depending on the specific context, these heuristics could also increase individual decision-competence. Moreover, (S)ocial facilitation (and maybe even (A)cceptance) are connected to the idea of *informational cascading*: the phenomenon that a group may end up taking a higher risk than any of its individuals would take, given the sequential influence of group members on each other.

Importantly, there are insights from social choice theory into how to mitigate against some of these potential traps in group decision making. We already mentioned the need to reflect critically on the weight and relevance of the evidence. A further suggestion for best practice is that after a successful reflection phase, each group member should submit their own judgment without any form of prior deliberation with others, anonymously and non-sequentially. Such an initial vote will help 'anchor' individuals to their own views without prior group influence and the potential for a cascading effect.

How can the vote be made anonymously and non-sequentially in an outdoor sports context? This could simply involve a deck of cards (a red card indicates not-safe, a black one safe), or otherwise colour-coded cards (green, yellow, red) when using scoring that are submitted to one group member and anonymously shared with the group. If that is too complicated, groups can at least decide to vote *non-sequentially*: group members share their initial decision at the same time (on three: thumbs up or thumbs down). If the initial decision is unanimous, the next two stages may not always be necessary.

Stage 4: open deliberation

Our main lesson from social choice theory for an effective deliberation phase is three-fold: first, it should focus on the agreed upon decision problem. Second, it should aim to increase decision-competence of the individuals, and third, it should do so while preserving independence of the individual judgments.

So our recommendation for best practice is that the deliberation phase should start with a reminder to the group of just what it is that is to be decided. Deliberation itself should, if, say, the decision problem concerns the stability of a slope, focus on differences in the assessment of the shared body of evidence. So, for example, the focus could be on sharing and aiming to clarify the identified known unknowns, or on identifying weaknesses in reasoning and evaluation of the evidence, or on identifying incoherent aspects of the evidence. Lastly, deliberation should be pursued in a way that preserves independence and this could be done by not disclosing the individual's initial vote, and by making sure that each member of the group has a say and by not trying to influence the deliberation phase by irrelevant features.

Stage 5: group decision

Once the deliberation phase is finished, individuals should again vote and do so anonymously and non-sequentially. Now, given the original framing of the decision problem, the group should use either a majority or a median vote to arrive at a group decision. As discussed in section 2, the majority or median vote is from an *evidential* perspective the best choice. Under favourable conditions, it is more accurate than the individual decisions.

Now, our discussion so far has mainly focused on factual matters about snow stability and how groups can make better decisions. We mustn't forget though that risky decisions also raise *ethical* issues. For example, in high stakes scenarios, it may simply be wrong–ethically speaking–to force some group members to commit to a high stakes ski descent just because the majority of the group decided it is a (sufficiently) safe choice. Hence, there may, in high stakes contexts, be ethical or practical reasons not to use a majority rule but instead a veto rule: if one member of the group opposes a risky choice, then the group opposes it. So, while from an evidential perspective our discussion suggests that we should use a majority or median rule to improve group decision making, we have to acknowledge that, in some contexts, ethical or practical concerns may count against using these rules.

5. Conclusions: lessons for avalanche education and outdoor risk management

The aim of this article is to bring social choice theory to bear on group decision making in avalanche terrain. In doing so, we are focusing on often-ignored social aspects in avalanche decision making. At the core of our discussion is the notion of decision-competence and the idea that individuals not only have to acquire the relevant skills taught in avalanche classes but also need to learn to reflect about their competence in a specific case. As such, our discussion is not aimed at replacing existing lessons in avalanche education. We propose to add a further layer to decision making that is (in parts) grounded in hard-earned knowledge in basic snow and avalanche science.

Given this, our lessons are best aimed at advanced or expert decision makers and groups who have the relevant expertise in snow science. Only when decision makers have the ability to make competent decisions can there be a worthwhile engagement on the reflective level. This reflection will, we believe, often have to be combined with a realistic understanding of ones own *fallibility*: the acceptance that in avalanche terrain even a highly experienced and trained decision maker cannot always and easily know that they know in a specific case. Sometimes, we can come to know more by accepting that we know less.

Having said this, even for introductory avalanche awareness classes or ski-touring classes more generally this discussion can have important educational and social benefits for the learning environment. Our lessons for group decision making encourage a more democratic approach that provides the framework in which each group member can gain autonomy in the learning process, and it empowers each individual of a group to have their own voice heard. It can also help to create a shared sense of group identity by reinforcing that each member's voice is important and counts. A more formal group decision tool can, in turn, also help educators and teachers to reduce the influence of an overly dominant group member. Finally, it's worth remembering that this decision making 'toolkit' is not restricted to high stakes scenarios. For example, during the planning phase, our group decision approach can be used to decide on how to plan a route, or where to locate so-called key places. During the day, it can be used to decide whether to even dig a snow pit, and if so where; or it can be used to decide which of several routes identified in planning is the preferred group option. Ideally, forming such group decisions by voting can be seen as a kind of a game that involves all group members, which may make the educational experience more enjoyable and group decisions more inclusive.

Our approach to group decisions can also be important in an operational avalanche context. As we noted earlier, the input votes need not be issued by individuals but may simply be results of different stability tests. These can be aggregated using either a majority vote, or when the outputs of such tests are non-binary (safe, inconclusive, not-safe) a median approach may be used. Alternatively, consider the context of avalanche forecasters who have to decide which level of avalanche warning to issue for a given region. Again our staged group approach can help shape an informed group expert decision. Given that the avalanche hazard level ranges from 1 (low) to 5 (very high), a scoring vote with the median as the group's choice will be the recommended method. Similarly, it can be used by a group of ski patrollers deciding on whether to open a specific slope or whether to adopt various avalanche mitigation methods before opening slopes or roads. Naturally, if decisions involve high stakes, expert groups may reasonably decide to adopt a veto rule so as to err on the side of caution. A consequence of this approach to expert decisions is that it helps flatten the 'command structure' and may help to avoid status bias when differently ranked professionals are involved in group decisions. Alternatively by using weighted votes an expert ranking could be introduced to the voting to better represent the variations in general competence or to avoid ties.

Given the wide range of examples, it should now be easy to see how our approach to group decisions can be made relevant in other kinds of outdoor and adventure sports. So, for example, whitewater or sea-kayaking often involve group decisions in the planning phase or about specific high-stakes situations. Provided that appropriate decision problems can be identified for other such sports, our discussion about group decision making can be equally useful in many other activities. Indeed, it may also be relevant in the context of coaching. It can help to combine advice from multiple expert coaches into a collective coach decision, or a coach may use our method to make sure that the athletes' voices are heard and included into the decision process.

Ultimately, our discussion aims to improve group decision making by identifying the conditions under which, and the mechanisms by which, groups can be more accurate than their individual members. We can't emphasise enough, though, that the machinery of collective intelligence is reversible. Under favourable circumstances there is added knowledge and safety in group decision making. Under unfavourable circumstances the machinery shifts without warning into reverse, exposing the group to an added risk. Now, it is important to remember that even the most rational decision may, in the end, not be the 'correct' decision, that is, even the most skilled and competent decision maker can get it wrong; but the more competent a decision, the less likely that is. Avalanche decision making is difficult, especially given the limited resources available, but there is an often untouched potential in the wisdom of groups: it's up to individuals to learn when and how to make use of it.

Notes

1. Alternatively, depending on the case, the right option could be the slope that would be the most fun to ski down, or the quickest, or that will take them back to their starting point, or whatever else. It doesn't matter *what* the rightness of this option consists in, only that this is independent of what the group happens to decide; in this limited sense, the rightness could be called 'objective'.
2. Note that we will say more about the 'favourable conditions' that are required to generate a wisdom of crowd effect. In particular, we will explain why judgements have to be *independent* in section 3. Note that the collective decision-competence of the group can be greater than the individual decision-competence of its most competent member as well, so that deciding together is better than picking a leader and letting them decide.
3. For more details about ECT's, see Simenhois and Birkeland (2009). For a recent discussion about potential cognitive pitfalls in interpreting such tests, see Ebert (2019).
4. For discussion about the binary output interpretation, the relevant threshold and their reliability, see Techel, Winkler, Walcher, Van Herwijnen, and Schweizer (2020).
5. Although this group happens to have three members, this decision method, like majority voting, is available for any group with an odd number of members: 3, 5 and so on. With an even-numbered group there are two judgments in the middle, when all of them are put in order, and some method for resolving differences is needed for those cases in which the two middlemost judgments disagree. Notice that this decision method generalizes the '2 + 1' decision design for three group members, in which the collective grade is the one that two group members agree on, if there is such an agreement, but otherwise the third person is consulted and the collective grade is the median of the three judgments. (Gottlieb & Hussain, 2015).
6. The reason for this is that if the collective score for one option is higher than that for a second option, and that is in turn higher than the collective score for a third option, then the collective score for the first option necessarily is higher also than the collective score for the third option. That follows immediately from the fact that the scores come in a 'top to bottom' order. Note, however, that while the decision method always produces coherent results it doesn't rule out the need for a tie breaker, i.e. different slopes may receive the same median score.
7. There are numerous existing academic debates about the independence assumption. See appendix B for further discussion and references.
8. Of course, a jury in a criminal trial is not usually adopting the majority rule, but rather juries in a criminal trial have to come to a unanimous decision. In the next section, we discuss a veto rule for group decisions.
9. A study by Waters and Hans (2009), involving 3500 jurors using the veto rule, found evidence for such a *reputation effect*: more than one in three juries contained at least one juror who privately would have voted *against* the jury judgment.
10. For a general overview of existing empirical studies which highlights potential pitfalls of and recommendations for group decision making in a variety of contexts, see e.g. Eller and Frey (2018). For what it is worth: some of the recommendations based on these empirical studies are closely related to our recommendations derived from the conceptual framework of social choice theory (compare also fn 11).

11. There are a number of different so-called structured group decision making tools used and studied in a variety of contexts such as operational research, economics, medical assessments, and forecasting. The most famous is the Delphi-method invented by the RAND corporation in the late 1940's and introduced as a method to improve the accuracy of forecasts. Here is not the place to discuss these decision tools in detail though some of our suggestions for best practice are influenced by them. For an extensive surveys of the Delphi-method, see Woudenberg (1991) and de Loë, Melnychuk, Murray, and Plummer (2016) who also discuss difficulties of measuring the effectiveness of a decision method. For more critical voices and a recent comparison between Delphi and its variants, see Graefe and Armstrong (2011).
12. Journey planning with key places for decisions is an important aspect of the 'Be Avalanche Aware' curriculum widely used in the UK and supported by the Scottish Avalanche Information Service (Scottish Avalanche Information Services, 2021). BAA is similar to the 3 × 3 method used in Switzerland and first developed by Munter (1992).
13. There are other potential ambiguities when eliciting risk judgements. For example, how verbal probability notions, such as 'likely', are understood can vary between individuals and in other work one of us distinguished three different formal notions of risk, see that may be implicit in our intuitive thinking about risk, see Ebert, Smith, and Durbach (2020).
14. The checklist stands for (I)llness, (M)edication, (S)tress, (A)lcohol, (F)atigue, (E)motion (or sometimes with (E)ating which also includes hydration).
15. Similarly, passing the *IM SAFE* checklist obviously isn't sufficient for being competent with respect to flying a plane, though failing the *IM SAFE* checklist is good evidence that you are currently not competent to operate the plane, no matter how much training and experience you have.
16. Indeed given that the avalanche decision-context is considered a 'wicked' learning environment with an unreliable feedback mechanism, finding out whether one is a competent decision maker in a general sense can be very difficult and there is scope for an overconfidence bias (Ebert, 2015; Hogarth, 2001; Zweifel & Haegeli, 2014). The important issue of identifying and assessing general competence in avalanche decision making is a thriving research project (Groves & Varley, 2020; Hallandvik, Andresen, & Aadland, 2017; Landrø, Hetland, Engeset, & Pfuhl, 2020; Stewart-Patterson, 2015).
17. The checklist goes back to McCammon (2004, 2009) and is mentioned in most avalanche textbooks and courses. It stands for (F)amiliarity with terrain, (A)cceptance within a group, (C)onsistency means less flexibility in decision making, (E)xpert halo, (T)racks in a slope, (S)ocial facilitation. McCammon's research suggests that these are 'heuristic traps', though as we argue below, these heuristics need not be traps and whether they are depends on the specifics of the case. It is worth noting that FACETS is not the only checklist for avalanche decision making. So, for example, the SOCIAL-checklist (Zweifel, 2014) also focuses on the human aspect, while the Avaluator (Haegeli, McCammon, Jamieson, Israelson, & Statham, 2006), and the previously mentioned 3 × 3 checklist focus more on the physical and snow-science aspect of avalanche decision making.

Acknowledgments

We are grateful to the audience of the Adventure Sports Coaching Conference organised by Doug Cooper at Glenmore Lodge in December 2019 and members of the Varieties of Risk seminar for their thoughts and comments. Special thanks to Mike Austin, Mark Diggins, Bruce Goodlad, Lesley McKenna, Ian Sherrington, Paolo Savino and Andrew Sullivan for written comments on an earlier draft. Finally, we are very grateful to the anonymous referees for extensive and thoughtful comments that improved this paper. Philip Ebert's work was supported by the Royal Society of Edinburgh (62,345) and the Arts and Humanities Research Council AH/T002638/1, 'Varieties of Risk'. Michael Morreau's work was supported by the Nordic funding agency NordForsk, under Grant NORDFORSK 105061.

Disclosure statement

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

Funding

This work was supported by the Arts and Humanities Research Council [AH/T002638/1]; NordForsk [105061]; Royal Society of Edinburgh [62345_Ebert_2019 Arts and Humanities Small Grants].

Notes on contributors

Philip A. Ebert is professor of philosophy at the University of Stirling. His main research concerns normative and behavioural issues relating to risk, rationality, and decision making using theoretical and experimental methods. His spare time is spent mostly in the mountains to ski, climb, run, or mountain bike, and he is lucky enough to have spent numerous seasons in Chamonix. He is also a qualified Mountain Leader (summer) and Backcountry Ski Leader and is a member of the Snow and Avalanche Foundation of Scotland (SAFOS) risk communication group. He also acts as an academic advisor to the Scottish Avalanche Information Services (SAIS).

Michael Morreau is professor of philosophy at the Arctic University of Norway (UiT) and senior researcher at the Center for Applied Research (SNF) of the University of Bergen. After studies at the University of Amsterdam, he worked first in computational linguistics and then in philosophy at Tübingen University and at the University of Maryland, College Park. Much of his recent research is about collective decisions made using scores or grades. Combining theoretical analysis, computer simulation and behavioural studies, this work aims to bring about practical improvements in the design of electorates, expert panels and other collective decision-making bodies throughout society.

ORCID

Philip A. Ebert http://orcid.org/0000-0002-3730-0600
Michael Morreau http://orcid.org/0000-0002-9507-5376

References

Anderson, L., & Holt, C. (1997). Information cascades in the laboratory? *American Economic Review*, *87*, 847–862.
Bikhchandandi, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, *100*(5), 992–1026.
Branch, J. (2012). Snow fall? The avalanche at tunnel creek. New York, NY: The New York Time (December 24). Accessed 11th August 2021 Retrieved from https://www.nytimes.com/projects/2012/snow-fall/index.html
Condorcet, J. A. N. C. (1785). *Essai Sur L'application de L'analyse a la Probabilite des decisions rendues a la pluralite des voix*. Paris: De L'Imperimerie Royal.
de Loë, R. C., Melnychuk, N., Murray, D., & Plummer, R. (2016). Advancing the state of policy Delphi practice: A systematic review evaluating methodological evolution, innovation, and opportunities. *Technological Forecasting and Social Change*, *104*, 78–88.
Dietrich, F., & List, C. (2004). A model of jury decisions where all jurors have the same evidence. *Synthese*, *142*(2), 175–202.
Dietrich, F., & Spiekermann, K. (2013a). Epistemic democracy with defensible premises. *Economics and Philosophy*, *29*(1), 87–120.
Dietrich, F., & Spiekermann, K. (2013b). Independent opinions? On the causal foundations of belief formation and jury theorems. *Mind*, *122*(487), 655–685.
Dietrich, F. (2008). The premises of Condorcet's jury theorem are not simultaneously justified. *Episteme*, *5*(1), 56–73.
Dietrich, F, and Spiekermann, K. (2021). Jury Theorems, In E. N. Zalta Ed., The Stanford Encyclopedia of Philosophy. Winter 2021 Edition. https://plato.stanford.edu/archives/win2021/entries/jury-theorems/
Ebert, P. A., & Photopoulou, T. (2013). Bayes' beacon: Avalanche prediction, competence, and evidence for competence. Proceedings of the International Snow Science Workshop, Grenoble, France, 363–370.
Ebert, P. A., Smith, M., & Durbach, I. (2020). Varieties of Risk. *Philosophy and Phenomenological Research*, *101*(2), 432–455.
Ebert, P. A. (2015). Know your competence. Reflecting on competent decision making in avalanche terrain. *The Avalanche Review*, *33*(4), 22–23.
Ebert, P. A. (2019). Bayesian reasoning in avalanche terrain: A theoretical investigation. *Journal of Adventure Education and Outdoor Learning*, *19*(1), 84–85.
Eller, E., & Frey, D. (2018). The group effect: Social influences on risk identification, analysis, and decision making. In M. Raue, E. Lermer, and B. Streicher (Eds.), *Psychological perspectives on risk and risk analysis: Theory, models, and Applications* (pp. 131–150). Cham: Springer.
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.
Galton, F. (1907). Vox Populi. *Nature*, *85*(5), 450–451.
Goodin, R. E., & Spiekermann, K. (2018). *An epistemic theory of democracy*. Oxford: Oxford University Press.
Gottlieb, K., & Hussain, F. (2015). Voting for image scoring and assessment (VISA) theory and application of a 2+1 reader algorithm to improve accuracy of imaging endpoints in clinical trials. *BMC Medical Imaging*, *15*(1).
Graefe, A., & Armstrong, J. S. (2011). Comparing face-to-face meetings, nominal groups, delphi and prediction markets on an estimation task. *International Journal of Forecasting*, *27*(1), 183–195.

Grofman, B., Owen, G., & Feld, S. (1983). Thirteen theorems in search of the truth. *Theory and Decision*, *15*(3), 261–278.
Groves, M. R., & Varley, P. J. (2020). Critical mountaineering decisions: Technology, expertise and subjective risk in adventurous leisure. *Leisure Studies*, *39*(5), 706–720.
Haegeli, P., McCammon, I., Jamieson, B., Israelson, C., & Statham, G. (2006). The avaluator–A Canadian rule-based avalanche decision support tool for amateur recreationists. Proceedings of the International Snow Science Workshop 2006, Telluride, CO, USA, 254–263.
Hallandvik, L., Andresen, M. S., & Aadland, E. (2017). Decision making in avalanche terrain–how does assessment of terrain, reading of avalanche forecast and environmental observations differ by skiers' skill level? *Journal of Outdoor Recreation and Tourism*, *20*, 45–51.
Harvey, S., Winkler, K., Techel, F., & Marty, C. (2013). *Schnee und Lawinen in den Schweizer Alpen. Hydrologisches Jahr 2006/07*. Davos, Switzerland: WSL- Institut für Schnee- und Lawinenforschung SLF.
Hastie, R., Penrod, S., & Pennington, N. (1983). *Inside the jury*. Cambridge, MA: Harvard University Press.
Hogarth, R. M. (2001). *Educating Intuition*. Chicago, IL: University of Chicago Press.
Landrø, M., Hetland, A., Engeset, R. V., & Pfuhl, G. (2020). Avalanche decision making frameworks: Factors and methods used by experts. *Cold Regions Science and Technology*, *170*, 102897.
Leiter, A. M. (2011). The sense of snow – Individuals' perception of fatal avalanche events. *Journal of Environmental Psychology*, *31*(4), 361–372.
List, C. (2013). Social Choice Theory. In E. N. Zalta Ed., *The Stanford Encyclopedia of Philosophy*. Winter 2013 Edition. Metaphysics Research Lab, Stanford University. URL=https://plato.stanford.edu/archives/win2013/entries/social-choice
Marder, N. S. (1987). Gender dynamics and jury deliberations. *The Yale Law Journal*, *96*(3), 593–612.
Marengo, D., Monaci, M. G., & Miceli, C. (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.
McCammon, I. 2009. "Human factors in avalanche accidents: Evolution and interventions." Proceedings of the International Snow Science Workshop, Davos, 644–648.
McCammon, I. (2004). Heuristic traps in recreational avalanche accidents: Evidence and implications. *Avalanche News*, *68*, 42–50.
Morreau, M. (2021). Democracy without enlightenment: A jury theorem for evaluative voting. *Journal of Political Philosophy*, *29*(2), 188–210.
Munter, G. (1992). *Neue Lawinenkunde: Ein Leitfaden für die Praxis*. Bern, Switzerland: SAC.
Scottish Avalanche Information Services. (2021). Be avalanche aware. Accessed 11th August 2021. Retrieved from http://beaware.sais.gov.uk
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*(2–3), 210–216.
Stewart-Patterson, I. (2015). Measuring decision expertise in commercial ski guiding in a more meaningful way. *Journal of Outdoor Recreation and Tourism*, *13*, 44–48.
Sunstein, C. R. (2000). Deliberative trouble? Why groups go to extremes. *The Yale Law Journal*, *110*(1), 71–119.
Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few*. London: Abacus.
Techel, F., Winkler, K., Walcher, M., Van Herwijnen, A., & Schweizer, J. (2020). On snow stability interpretation of extended column test results. *Natural Hazards and Earth System Sciences*, *20*(7), 1941–1953.
Tremper, B. (2018). *Staying Alive in Avalanche Terrain* (3rd edition ed.). Seattle: Mountaineers Books.
Waters, N., & Hans, V. (2009). A jury of one: Opinion formation, conformity, and dissent on juries? *Journal of Empirical Legal Studies*, *6*(3), 513–540.
Woudenberg, F. (1991). An evaluation of Delphi. *Technological Forecasting and Social Change*, *40*(2), 131–150.
Zweifel, B., & Haegeli, P. (2014). A qualitative analysis of group formation, leadership and decision making in recreation groups travelling avalanche terrain. *Journal of Outdoor Recreation and Tourism*, *5-6*, 17–26.
Zweifel, B. (2014. SOCIAL–A group check tool. Proceedings of the International Snow Science Workshop, Banff, AB, Canada, 963–969.

Appendix A. Condorcet's Jury Theorem: visual representation

We provide a visual representation of the basic idea behind the Condorcet Jury Theorem. We do so by means of an analogy: suppose you have a heavily biased coin, such that there is a 2 in 3 chance that it gives you tails and 1 in 3 chance of heads. Assume that in this case the 'True' answer is 't'ails, so that flipping the coin is like having a 2 in 3 chance of getting the correct answer. Now, imagine, that you have three different coins that are each biased towards providing the 'true' answer and they are independent of each other. You are allowed to flip each one of them and you then chose the majority answer. In this case, the chance of getting the right answer after all three coins are flipped will be close to 3 in 4 (it increased from a 66.6% to a 74.1% chance).

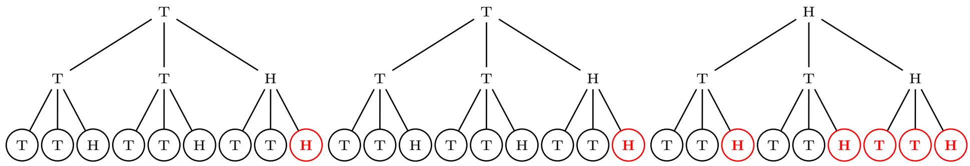

Figure A1. Visual representation of throwing a biased coin. The top row represents the first coin toss (each node is equi-probable), the middle row represents the second coin toss, and the bottom row represents the third coin toss. The circled outcomes in the third line are all the equi-probable outcomes of throwing three biased coins in an order.

Table A1. Table showing the accuracy of group judgment given prior individual decision-competence of a homogenous group with 3 or 5 independent voters using the majority rule.

Individual Competence	Group of 3 Competence	Group of 5 Competence
33.3%	∼ 25.9%	∼ 21%
50%	50%	50%
66.6%	∼ 74.1%	∼ 79%
80%	∼ 89.6%	∼ 94.2%
90%	∼ 97.2%	∼99.1%

To see why, consider Figure A1. The first line has three so-called *nodes*: T, T, H that each start a *tree*. Each node in the first line, represents a 1 in 3 chance of that outcome. Thus, the first line visually represents the first coin toss and a 2 in 3 chance of throwing tails (T). The second line represents the throw of the second biased coin, and the third line the throw of the third biased coin. The circled outcomes in the third line are all the equi-probable outcomes of throwing three biased coins in order. The black circled outcomes are those where a *majority vote* results in choosing tails, that is the true answer, the red (boldfaced) ones where the majority rule requires you to choose heads, that is the false answer. As a result, we get the true answer in 20 out of 27 (74%) cases when throwing three (independent) coins that each have a 66.6% chance of getting it right, and in 7 out of the 27 (26%) scenarios we opt for the false answer. Hence, a decision based on three such coin tosses is more reliable than on one coin toss.

From our perspective, we can interpret individual decision-competence as akin to a (possibly) biased coin. If three people are each more likely than not to get the right answer and are able to provide *independent* judgements, a wisdom of crowd effects arises for them just as for the three biased coins above. This visual representation also nicely explains, as emphasised in section 2, why and how the theorem can 'backfire': if individual decision-competence is say 33.3%, then aggregating amongst these decision makers leads to a reduction of group decision-competence to 25.9%. Table A1 collates other examples how aggregation can increase (or decrease) decision-competence depending on group size (3, 5 people) and different case-specific competence levels of individuals (33.3, 50, 66.6, 80, 90).

A corollary of CJT is that as the number of judgements increases, the accuracy increases with a limiting case of perfect accuracy given an infinite number of independent judgements (for an overview of this and other relevant formal results relating to CJT, see Grofman, Owen, and Feld (1983)). Naturally, it seems somewhat surprising to think that we could, even in principle, arrive at perfect accuracy in real-life settings, such as avalanche predictions. However, when the more realistic assumption is made that competent judgements are made relative to a body of evidence which may not be conclusive, the limiting case is not necessarily perfect accuracy. In appendix B, we outline a more realistic version that expands on this consideration.

Appendix B. Condorcet's Jury Theorem: best responder interpretation and fixed problem interpretation

The so-called best responder interpretation of CJT reinterprets the judgements that individuals make as conditional on a shared body of evidence–just as we have presented the situation in the main text. The best responder constitutes the most rational response to the evidence. Note that this interpretation makes some substantial assumptions. For example, that there is a unique best response to the evidence and that the relevant accuracy is point valued (rather than interval-valued). On this interpretation, a 'best responder' limit is introduced, and it is this limit to which aggregating judgements trend given sufficient decision-competence. Hence, rather than the unrealistic result that a large number of respondents will make their judgment nearly infallible, a large number of decision-competent respondents will simply converge on the judgment of the best responder and her decision-competence. Especially in the case of avalanche decision making where we don't operate with conclusive evidence but with remaining uncertainties within a body of evidence, a limit that in effect reflects the quality of the evidence, seems a more realistic interpretation. This interpretation is presented in a very accessible way in Goodin and Spiekermann (2018), which draws on formal discussion by Dietrich and List (2004); Dietrich and Spiekermann (2013a, 2013b, 2021).

Furthermore, in contrast to most appeals of CJT, we adopt what Dietrich (2008) calls a *fixed problem CJT*-interpretation as opposed to a *variable problem CJT* interpretation. In a fixed problem interpretation, the competence and independence conditions are fixed (i.e. conditionalised) to the specific case under considerations. This is how we use CJT throughout our discussion. It is a direct result of this interpretation that our main focus is the issue of decision-competence, since it is that condition, and not the independence condition, that is more difficult to meet. Our emphasis on the importance of rational reflection on ones decision-competence–the known knowns and known unknowns–in a given case, is part of our response to the challenges raised by Dietrich (2008) against the *fixed problem CJT*-interpretation.

Appendix C. Collective grading versus majority voting

In many instances, a group can arrive at a choice between two options in either of two ways. One way is collective grading. First, each member of the group grades both of the options; second, for each option, the individually assigned grades are aggregated in some suitable way, say by taking the median; finally, the group chooses the option with the highest (or best) collectively assigned grade. Another way for a group to decide is by holding a majority vote. Each member of the group votes for one of the options (the one that they personally think is best) and the group chooses whichever option receives the most votes.

Suppose one of the two options between which a group is going to choose is, in some antecedent sense, the right one for it to choose. The CJT identifies conditions on the decision-competence of individual group members under which majority voting is likely to result in the group's choosing this right option (provided individual vote decisions are suitably independent). The grading-jury theorem (Morreau, 2021) similarly identifies competence conditions under which median grading is likely to result in the right choice (again assuming independent individual decisions), and shows that sometimes the competence condition for median grading is satisfied while that for majority voting is not. When such circumstances obtain, the group would do better to frame the choice before them as a grading problem rather than as a majority vote, because this makes it more likely that they will make the right choice. Intuitively, these are cases in which biases pull group members in the wrong directions: people are inclined to give low grades to the right option and high grades to the wrong option. Now, we illustrate qualitatively how such biases might arise with a binary decision in avalanche terrain.

Suppose Anne, Bertil, and Carsten are deciding together whether to ski the northern or the western slope. They've agreed to choose the safest of the two and that happens to be the western slope. So, the western slope is the right one for them to choose. Anna, though, happens to be very familiar with the northern slope and has skied it quite often before without any problems. As a result, she perceives it as much safer than it really is, and communicates this conviction to the others, who don't have the same experience but are willing to take her word for it. Furthermore there are visible tracks in the northern slope, again making it seem much safer than it actually is. On the other hand, from where everybody is standing the western slope appears very steep, and so everyone in the group is inclined to think it more risky than it really is. Under such circumstances, biases and misleading perceptions pull everybody's judgments in the wrong directions. That is, everybody *under*estimates the risk of the *more* risky option and everybody *over*estimates the risk of the *less* risky option.

The result (depending on the strength of the biases) can be that each member individually is most likely to judge that the northern slope is safer than the western slope, and to vote for it, when really it is just the other way around, and they should vote instead for the western slope. In this case, the CJT tells us, holding a majority vote about which slope to ski tends to amplify the individual perceptions, and increases the risk of making the wrong choice. In some such cases, though, it can be shown, the competence condition of the grading-jury theorem still is satisfied. Section VI of (Morreau, 2021) discusses a precise example of this sort and presents results of computer simulations in which groups are more likely to make the right decision by collective grading than by majority voting.