Journal of Adventure Education and Outdoor Learning

ISSN: 1472-9679 (Print) 1754-0402 (Online) Journal homepage: http://www.tandfonline.com/loi/raol20

Bayesian reasoning in avalanche terrain: a theoretical investigation

Philip A. Ebert

JOURNAL OF ADVENTURE EDUCATION AND OUTDOOR LEARNING
2019, VOL. 19, NO. 1, 84–95
https://doi.org/10.1080/14729679.2018.1508356

Bayesian reasoning in avalanche terrain: a theoretical investigation

Philip A. Ebert

Division of Law and Philosophy, University of Stirling, Stirling, UK; Behavioural Science Centre, University of Stirling, Stirling, UK

ABSTRACT
In this article, I explore a Bayesian approach to avalanche decision-making. I motivate this perspective by highlighting a version of the base-rate fallacy and show that a similar pattern applies to decision-making in avalanche-terrain. I then draw out three theoretical lessons from adopting a Bayesian approach and discuss these lessons critically. Lastly, I highlight a number of challenges for avalanche educators when incorporating the Bayesian perspective in their curriculum.

KEYWORDS
Bayesian reasoning; avalanche decision-making; avalanche education; base-rate fallacy; snow science; risk management

1. Introduction

In recent years, there has been extensive discussion of the human factor in avalanche decision-making. Avalanche educators have pointed to various well-known biases and heuristic traps that may lead outdoor enthusiasts to misjudge the risks when making decisions in avalanche terrain (McCammon, 2004; Furman, Shooter, & Schumann, 2010; Leiter, 2011; Marengo, Monaci, & Miceli, 2017; and others). While much emphasis has been placed on a number of heuristic traps highlighted in McCammon (2004), such as the 'expert halo', the 'role of familiarity' or the 'herd instinct', little attention has been paid to a well-known *cognitive* bias concerning reasoning and thinking about risk called the 'base-rate fallacy' or sometimes labelled 'base-rate neglect'. When committing this fallacy, intuitive thinking involving probabilities will be systematically off-target and can lead to serious misjudgements about the risk of certain events. Furthermore, this kind of neglect seems wide-ranging: it is a phenomenon widely studied in psychology, philosophy, and neuroscience, with applications in legal theory, medical sciences, and economics, affecting subjects with different educational backgrounds (e.g. Kahneman & Tversky, 1973, Bar-Hillel, 1980, Barbey & Sloan, 2007, Pennycook & Thompson, 2017. See Birnbaum, 2004 and Gigerenzer & Hoffrage, 1995 for more critical discussions.) The aim here is not, however, to investigate whether backcountry skiers in particular are more or less prone to this particular fallacy, rather the article is more *theoretical* in nature: using the base-rate fallacy, I wish to motivate a broadly Bayesian perspective on avalanche decision-making and outline some of the more theoretical consequences of adopting Bayes' theorem—the theorem that underlies the Base rate fallacy—more seriously in an avalanche decision-making context.

The article is structured as follows: first, I will offer a brief outline of the base-rate fallacy and Bayesian reasoning more generally and then show how this approach can be made relevant in the context of avalanche decision-making. Based on these considerations, I offer three general lessons for avalanche decision-making from a Bayesian perspective. I will then assess three challenges to

CONTACT Philip A. Ebert p.a.ebert@stir.ac.uk

this perspective, and, in the last section, initiate a discussion about the value and difficulties of the Bayesian approach for avalanche education more generally.

2. Bayesian reasoning and the base-rate fallacy: a quick introduction

The base rate fallacy became prominent through the work of Kahneman and Tversky (1973). It occurs when a subject does not take the so-called base rate information seriously and reasons in a way that goes against a theorem of probability called *Bayes' theorem*. Let us briefly outline one of the classic examples, inspired by Eddy (1982) as discussed in Gigerenzer and Hoffrage (1995) which is referred to as the *Mammography problem*:

> The probability of breast cancer is 1% for a woman at age 40 who participates in routine screening. If a woman has breast cancer, the probability is 80% that she will get a positive mammography (hit rate). If a woman does not have breast cancer, the probability is 9.6% (false positive) that she will also get a positive mammography. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

According to Eddy's (1982) informal sample 95 out of 100 physicians misinterpreted the statement about the accuracy and estimated the probability of breast cancer, given a positive diagnostic test, to be somewhere between 70% and 80%. Since then many more formal studies have been conducted using different contexts and the results are (broadly speaking) similar (Casscells, Schoenberger, & Grayboys (1978), for more variation in the results, see Gigerenzer & Hoffrage (1995)). The correct answer in the above case is merely 8% and thus much lower than the intuitive estimates. To see how to arrive at this result, we can either draw on Bayes' theorem explicitly (for details, see appendix A) or, more simply, we can use the following frequency reasoning: consider 1000 women of whom 10 will have breast cancer (1% base rate). Given that the hit rate is 80% the test will come back positive for 8 out of 10 women with breast cancer. However, given the false-positive rate is 9.6%, roughly 95 of 990 women without breast cancer will get a (false) positive result. Hence the chance to actually have breast cancer given a positive result is 8 out of a total of 103 (95 + 8) women, so just under 8%.

The fallacy arises when subjects neglect to give the base rate information-that only 1% of woman at the age of 40 have breast cancer-its proper weight and end up overestimating the implications of a positive test result.[1] As a result, (Eddy, 1982) notes that 'physicians do not manage uncertainty very well, that many physicians make major errors in probabilistic reasoning, and that these errors threaten the quality of medical care' (Eddy, 1982, p. 249). The issue I want to raise is whether there is a similar *potential* for 'major errors' when decision-makers use stability tests in avalanche terrain. After all, avalanches are fairly rare events (Pfeifer, 2009, p. 428) and slope stability tests are known to be not perfectly accurate (Tremper, 2008).

In order to motivate Bayesian reasoning in an avalanche decision-making context, we first require the base rate information. For simplicity, let's say that a slope is *not-safe* if and only if a skier would trigger a decent sized avalanche greater than size 1 on that slope (for sizing of avalanches see McClung & Schaerer 1993). So the required base rate will be the probability that a slope is not-safe, *prior to having applied slope specific diagnostic tests*. This type of data is difficult if not impossible to acquire without knowing more about the number of people engaged in the activity, the time spent in avalanche terrain, and most importantly the competencies of the skiers involved (Techel, Zweifel, & Winkler, 2015; Winkler, Fischer, & Techel, 2016). Nonetheless, some relevant information is available, which is sufficient to show how Bayesian reasoning can have a great potential in avalanche decision-making.

According to Zweifel et al. (2006)—as presented in Jamieson et al. (2009)—the avalanche fatality rate per day is roughly 1:70,000 in an area around Davos, Switzerland. More recently, Winkler et al. (2016) suggest a slightly lower fatality rate of just below 1:100,000 for ski touring in Switzerland. Using a fatality rate of roughly 1 in 10 people who are caught in an avalanche of decent size

(Schweizer & Lutschg 2001; Brugger et al. 2007; Jamieson et al. 2009, highlighting variation), we arrive at a rate of triggering avalanche of around 1:7,000 (0.014%) to 1:10,000 (0.01%) per day per skier in that area. In contrast, Grímsdóttir and McClung (2006) offer more worrying statistics. They recorded 345 avalanches for over 200,000 ski runs in a certain are of Canada (roughly 17 per 1000 runs). This data gives us roughly a 1.7% chance of triggering a decent sized avalanche per run. Unfortunately, in all this data little is known about the competences of people involved and to what extent skiers actually employed any slope specific stability tests. Given this variability, let's assume for our purposes a risk that is at the higher end, such as a 1 in 100 chance that on a given run an avalanche will occur *prior* to applying any stability test. We will deal more specifically with the inherent uncertainty and, importantly, the variability of our chosen base-rate information in Sections 4.2 and 4.3.

Second, we also require information about the accuracy of diagnostic tests for slope stability, in particular, we require the 'hit rate' and the 'false-positive rate'. Again, it is very difficult to acquire reliable data (Simenhois & Birkeland, 2009; Birkeland & Chabot, 2006). But what is clear is that there is no fool-proof diagnostic test to decide whether a given slope is *safe* or that it is *not-safe* (Tremper, 2008). Moreover, any given test which is currently used is subject to much interpretation and errors can be introduced simply through human fallibility or non-representative sampling (Hendrikx & Birkeland, 2008; Hendrikx, Birkeland, & Clarke, 2009). Studies suggest hit rates for different tests to be between 56% and 94% and false-positive rates to be between 0% and 18% (Simenhois & Birkeland, 2009; Gauthier & Jamieson, 2008). Let's assume for simplicity (complications will be discussed in Section 4.2) that for a broadly competent decision-maker the relevant diagnostic tests have a 80% hit rate (so 8 out of 10 not-safe slopes will be identified as not-safe) and that there is roughly a 10% false-positive rate (1 out of 10 safe slopes are identified as not-safe).

Using these ballpark figures, we can now put together a similar template, which we label the *Avalanche Problem*:

> The probability of a skier triggering a decent sized avalanche on any given skiable slope in the area is 1%. If a slope is not-safe, a stability test will have 80% of indicating that it is not-safe (hit rate). If a slope is safe, the probability that the stability test indicates it is not-safe is 10% (false-positive). A stability test is applied to a skiable slope with the result indicating that it is not-safe. What is the probability that the slope is not-safe?

And, unsurprisingly, the probability that a slope is *not-safe*, provided the test indicates it is *not-safe*, is 8%. As a result, in only 1 in 13 cases when the stability test indicates that a slope is not-safe, will there be an avalanche (given the above guesstimates for base rate and accuracy of a test). In what follows, we will discuss how this *thought experiment* may be of value for avalanche decision-makers and in avalanche education.

3. Three lessons of Bayesian approach to avalanche decision-making

Our intention here is not to test explicitly whether those navigating through avalanche terrain are more or less susceptible to this kind of base rate neglect. Given the wide-ranging studies that have been conducted by many different researchers in various different fields, using different kinds of questions and set-ups, it would be surprising if it were not the case that a sizeable proportion of this group of subjects are susceptible to a similar fallacy.[2] Rather, the above considerations are used to highlight how Bayesian thinking *could* be implemented in avalanche decision-making more generally by drawing out three lessons for avalanche decision-making.

One natural way to incorporate the Bayesian approach to avalanche decision-making is to regard the general avalanche guidance issued by local authorities as an indication of the relevant base rate.[3] Avalanche forecast authorities worldwide provide guidance using a standardized five point scale: level 1. is described as *low* risk; level 2. is *moderate* risk; level 3. amounts to a *considerable* risk; level 4 constitutes a *high* risk; level 5. stands for *extreme* risk. Recent research confirms that such general guidance does indeed track the underlying risk of triggering an avalanche (Techel et al., 2015). The proportionality between guidance and risk increase is hard to quantify: a one level increase in the general avalanche guidance increases the risk of triggering an avalanche by a factor somewhere between 2 and 10. Pfeifer (2009) and Munter (2003) suggest a factor close to 2, Techel et al. (2015) suggest factor 2.6–2.7, Jamieson et al. (2009) suggest factor 10. It might well be that the increase is non-linear.

If, for the moment, we assume that an assessment level of *considerable* reflects the above used base rate of 1 in 100 and we adopt a factor of 4 for an increase or decrease of one level of the general avalanche warning, then an avalanche warning of *high* would correspond to 1 in 25, while a moderate danger amounts to 1 in 400 chance of triggering a decent sized avalanche prior to applying any stability tests. With these base rates in place, we can highlight an important aspect of how variations in base rate information may change the 'meaning' of our stability tests.[4] Consider a moderate warning scenario where the base rate is 1 in 400. A lower base rate lowers the probability of a slope being not-safe, given a stability test result that indicates that it is not-safe, to less than 2%. So, if the base rate is very low–as can be assumed on moderate avalanche warning days–the test's so-called *positive predictive value* is low and only 1 in 50 slopes that are tested not-safe, will indeed be not-safe! Crucially, the converse is true as well: in higher avalanche danger, e.g. if the base rate is 4% (i.e. 1 in 25), then the probability that a slope is not-safe given that the diagnostic test predicts it is not-safe is 20%, i.e. 1 in 5 slopes that test not-safe are indeed not-safe. So, in higher avalanche danger, the predictive value of the test is *substantially* higher and stability tests are more likely to be correctly predicting a slope's instability.

This consideration constitutes the first important lessons of the Bayesian approach to avalanche decision-making:

First lesson of the Bayesian approach: The role of base rates
Localised diagnostic test will be more *informative* the higher the general avalanche warning.

Of course the flipside is that diagnostic tests become less informative in moderate and low avalanche terrain. Yet, again we think this insight is important and contains another important lesson for avalanche decision-makers with regards to feedback and updating which is generally under-appreciated. To highlight the issue consider the following scenario:

Imagine a competent decision-maker, call her *Christelle*, skiing during moderate avalanche danger. Assume, she is subject to the above base rate neglect. Having taken an avalanche course she thinks that the diagnostic tests are not merely fairly reliable but also that they are highly informative (i.e. she thinks that the probability that a slope is not-safe given the test predicts it is not-safe, is very high). Moreover, imagine that Christelle is a responsible decision-maker and adheres to the result of her tests and so never skis the relevant terrain that tests not-safe. Unfortunately, however, and as many skiers will have experienced, she skis in an area that is home to numerous much more risk-seeking skiers. The likely feedback that Christelle will receive on moderate avalanche danger days is that most often, 'reckless' skiers who ski a slope that previously tested *not-safe* actually turns out to be *safe*. Thus the *feedback* she receives seems to undermine her test results. A very natural response to that sort of feedback would be for Christelle to question the value of her diagnostic test and treat it as inaccurate, unreliable, or simply not fit for purpose. Coupled with the assumption that Christelle is unaware of our first lesson, such a response may be disastrous and could, in high avalanche danger, make it very likely for her to trigger an avalanche.

What this consideration highlights is that an understanding of how diagnostic tests work, the base rate neglect, and more broadly the Bayesian way of thinking, will offer educators the tools to forewarn students not to draw the wrong 'lessons' from 'wrong' tests. Hence:

Second lesson of the Bayesian approach: The problem of misleading feedback

Avalanche terrain is a 'wicked' learning environment and does not reliably behave as predicted (in particular if the stability test predicts not-safe). Hence, do not 'blame' the stability tests for false positive results: they are to be *expected* when the avalanche danger is low. In fact, their existence is a consequence of the basic fact that low-probability events are difficult to detect reliably.[5]

Lastly, it is important not only to look at scenarios when a test returns a not-safe verdict but also when they return a safe verdict. This will help to emphasise the role of diagnostic test as a *method for risk reduction*, and not as a *method to settle with any high degree of confidence the status of a slope*. Remember that, in a moderate environment given our assumptions, a non-safe test means that a slope only has a 2% chance of being not-safe (i.e. 98% chance of being safe). A 1 in 50 chance of triggering an avalanche of decent size is, however, very risky. Yet, if we think of tests more generally as risk reduction tools, we need to put the 1 in 50 risk into a wider context: after all, what if a different slope actually does test safe? Assuming for simplicity, the same hit rate and false-positive rate as above, the probability that a slope is safe, given the test predicts it is safe, is 99.97%. Hence we arrive at our last insight:

Third lesson of the Bayesian approach: The method of risk reduction[6]
In avalanche decision-making, there is no certainty, all we can do is to apply tests to reduce the risk of a bad outcome, yet there will always be a residual risk.

In what follows, we will consider a number of challenges and clarifications to the Bayesian approach to avalanche decision-making and assess to what extent the main lessons are dependent on various guesstimates we adopted in this theoretical exercise.

4. Discussion and challenges

4.1. *The problem of oversimplification*

One way to challenge this theoretical exercise is to highlight that avalanche decision making is much more complex than we make it seem. Competent decision-makers continually look for ‘red flags’, they make assessments throughout the day and make general observations while approaching their intended slope. So, in combination with a carefully chosen snow pit, a decision maker may then perform a range of different stability tests (compression test, shear test, extended column test, etc.) to come to an informed decision based on a whole range of observations. Hence, our characterisation seriously *oversimplifies* the way competent decision-makers tend to make judgements in avalanche terrain.

Of course, there is no denying that this description is a much a more realistic characterisation of how *competent* decision-makers arrive at their judgements. However, these simplifications need not affect the considerations that were offered in Section 3. After all, only in so far our simplifications lead to a serious *underestimation* of the accuracy of the diagnostic test or tests, does this challenge have the potential to threaten some of our suggestions. We simplified by assuming that the stability test adopted will have a hit rate of 80% and a false-positive rate of 10%. However, whether we consider all tests together or consider for simplicity just one test makes no difference to the calculation so long as the former tests taken together or the latter individually have the same overall accuracy. Hence, in so far as the problem of oversimplification raises issues with the presumed accuracy of the diagnostic test, it will be better to consider this type of concern under the label of the next challenge: the problem of guesstimates.

4.2. *The problem of guesstimates*

Before tackling the problem of using guesstimates and ballpark numbers directly, it is worth considering the resulting risks in a wider context. Using the guesstimates adopted earlier, and assuming that you are a competent decision-maker, and that you only ski slopes that test ‘safe’ on a moderate avalanche warning day, there will be a 0.03% probability that you trigger a decent-sized avalanche on a given run. Let us assume that a skier skis three different (independent) runs a day so that a competent decision-maker faces roughly a 0.1% chance of triggering an avalanche per day ski touring on a moderate avalanche level. That is, she has a 1 in 1000 chance that day to trigger an avalanche, and given the assumed fatality rate of 1 in 10, she faces a 1 in 10,000 chance, or 100 in 1 million chance of dying in an avalanche per ski day (on a moderate warning day). Risk analysts often use the concept of a *micromort* to characterise the dangerousness involved in a given activity (Howard, 1980; see also Spiegelhalder & Blastland, 2014 which contains a chapter on extreme sports). One micromort is defined as a 1 in 1,000,000 chances of dying. So adopting our guesstimates and the micromort scale, a ski tourer who is a competent decision-maker 'uses' 100 micromort per day when engaging in this activity on a moderate avalanche rating.[7]

Is this realistic? Well, let us compare this risk with other activities: so, for example, skiing within a ski resort involves roughly 1 micromort (https://www.nsaa.org/media/68045/NSAA-Facts-About-Skiing-Snowboarding-Safety-10-1-12.pdf), skydiving involves roughly 10 Micromort (http://www.bpa.org.uk/staysafe/how-safe/), while doing a competitive marathon race incurs 7 micromort (Kipps, Sharma, & Pedoe, 2011). Base jumping, usually perceived as a very dangerous activity, requires when performed at the Kjerag massif in Norway roughly 432 micromort (Soreide, Ellingsen, & Knutson, 2007). It is difficult to tell, but maybe the fatality risk (by triggering an avalanche only) of a skier and competent decision-maker in avalanche terrain on a moderate avalanche level warning, is a little high given this wider context.[8] If so, this suggests that we might have overestimated the base rate information or underestimated the accuracy of our tests. Let's briefly consider these options in turn and see how this affects our three lessons.

If we lower our base rate for a moderate danger day, our three lessons for avalanche decision-making from a Bayesian perspective remain relevant. The first lesson will remain correct so long as an increase in avalanche rating corresponds to an increase in the base rate: a lower base rate will lead to a reduction of the informativeness of our tests. Moreover, lowering the base rate will also lower the probability of skiing a not-safe slope given a test predicting it is not-safe. Hence, lowering the base rate will make the problem of feedback alluded to in lesson two even more relevant. Lastly, lesson three remains applicable as well. While a lower base rate further decreases the residual risk, there will always be a residual risk. Hence, lowering the base rate for each avalanche warning does not significantly affect (and in part even increases) the relevance of our three lessons.[9]

Alternatively, and as alluded to in Section 4.1 our guesstimates about the accuracy of our test may be off-target. What if we underestimated both the hit rate as well as the false-positive rate of the various tests taken together? According to (Simenhois & Birkeland, 2009) certain types of test have a hit rate as high as 94%. To see how this affects our discussion let us increase the rate from 80% to 95% and reduce the false-positive rate from 10% to 5%. This affects the relevant risks (Table 1; Figure 1) as follows:

Assessing the three lessons above in light of these results, we easily see that lesson one is still supported: the higher the avalanche rating the greater the *informativeness* of the test. Moreover, we can see that an increase in accuracy of the tests, results in an increase of informativeness (all else equal). So lesson 1 remains correct when we increase the accuracy of our tests.

Lesson two contains a warning about feedback and highlights that given a low base rate, it will be more likely than not that a slope is safe, even if the test results indicate that the slop is not-safe. Again, this observation is borne out even with a more accurate test. On a moderate and

Table 1. The effect of increasing accuracy of tests on the posterior risks: the left column contains the base rate which is by assumption fixed by the general avalanche warning. The other two columns present the probability that a slope is not-safe given a not-safe test result. Middle columns use a lower accuracy (80% hit rate; 10% false-positive rate), while the right-hand column shows the results using a more accurate test (95% hit rate; 5% false-positive rate).

Avalanche rating (Scale; Base Rate)	[80–10 accuracy]	[95–5 accuracy]
Low; 1/1600 (0.063%)	~5/1000 (0.5%)	~12/1000 (1.2%)
Moderate; 1/400 (0.25%)	~2/100 (1.9%)	~5/100 (4.5%)
Considerable; 1/100 (1%)	~7/100 (7.5%)	~16/100 (16.1%)
High; 1/25 (4%)	25/100 (25.0%)	~44/100 (44.2%)
Extreme 1/6 (16%)	~62/100 (61.5%)	~79/100 (79.2%)

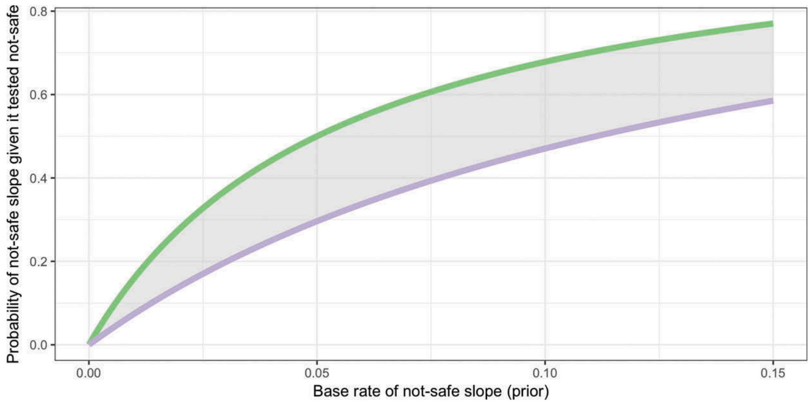

Figure 1. The effects of increasing accuracy of tests on the posterior risk. Purple line represents lower accuracy (80% hit rate; 10% false-positive), while green line represents higher accuracy (95% hit rate; 5% false-positive). Grey area represents variability of posterior given variation in accuracy and base rate.

considerable rating using the more accurate test, the probability that a slope is not-safe given a not-safe test is roughly 4.5% and 16.1%, respectively. Hence, there is still a danger to misinterpret the most likely outcome (no avalanche) as evidence that the test is unreliable, which is what lesson two warns against.

Lesson three also remains intact and the variation of the accuracy of the test nicely highlights how effective avalanche training can be to reduce the risk. Only if our test is maximally accurate will there be no residual risk–a scenario which is extremely (to say the least) unlikely.[10]

Let me make one last observation about the effects of increasing the accuracy of the relevant test. If we assume that the higher the *competency* of an avalanche decision-maker, the greater the *accuracy* of their stability tests (these tests are subject to interpretation and thus their accuracy depend on a subject's skill and experience), and we assume a skier with *no* competence to have a 1 in 100 chance of getting in an avalanche on a moderate warning per day, what given our assumptions would the relevant probability be for a competent decision-maker? That is, what is the probability that she skis a not-safe slope, given that she avoids all slopes that test not-safe? Assuming that a test with the lower accuracy rate represents a subject with some degree of competence, such subject reduces the risk from 100 in 10,000 (1%) to 22 in 10,000 (0.22%), i.e competence reduces the risk by factor 4.5. Assuming the more accurate test represents an even more competent decision-maker, such a decision-maker will reduce the risk to 5 in 10,000 (0.05%). Hence, competence would reduce her risk by factor 20, assuming, of course, all else remains equal, in particular a subject's risk attitude.[11]

4.3. *The problem of variability*

Appreciating the great variability of the risk of triggering an avalanche might give rise to one last concern. To put it simply: risk-averse skiers, who ski in an area that they know incredibly well on gentle slopes, will incur a very different risk than risk-seeking skiers, who ski new terrain on steeper slopes even if the general avalanche level that day is the same. So what's the point of making such general calculations?

Of course, we have to acknowledge that given these many factors there is a limit as to how specific we can be about an individual's risk. However, this is a general problem that affects statistical information in general.[12] Moreover, the greater the variability of the risk, the lower the relevance of the mean average when considering an individual's risks. And so, while we need to be very careful when applying the above risk statements to a specific case, there is still value in considering the above scenarios as a theoretical exercise since the three lessons will not be affected by this variability.

Nonetheless, to alleviate aspects of this concern, we can go one step further and deal with the variability directly and explicitly model it quite simply like this: assume that we say that a moderate day has a base rate of anywhere between 1/400 to 1/100 chance of triggering an avalanche and that the accuracy of the person applying the test is anywhere between 80–10 to 95–5, then we can say that the risk of triggering an avalanche on a slope that is identified as not-safe is in the interval from 1.9% to 16.1%. In effect, we can use Figure 1 to highlight just that variability: the grey area represents the relevant variability given changes in accuracy (purple-green line) and base rate (x-axis). Of course, we could go further and model that variability in better ways by assigning much more informative priors to the prior distribution of the accuracy rate instead of using, as I did here, a uniform distribution between the low and higher accuracy rate.[13] Importantly, however, making the variability more explicit or even modelling it in more sophisticated ways does not undermine the spirit of the main three lessons of a Bayesian approach to avalanche decision-making.

5. Conclusion: the challenge for avalanche educations

The theoretical exercise of using Bayesian reasoning contains a number of lessons about reasoning in avalanche terrain that are not widely acknowledged—the exception being the more technical discussion in McClung (2011). I argued that they are relevant despite the variability and uncertainty of the data. Given that Bayesian reasoning is, currently, one of the best tools we have to reason with uncertainties and probabilities, and given that avalanche decision-making is, in effect, decision-making under uncertainty with high stakes, it seems prudent to take Bayesian reasoning seriously in this context.

Nonetheless, the above considerations also represent a huge challenge to avalanche educators: Bayesian reasoning is not very intuitive (hence the fallacy alluded to above) and it is easily misunderstood. So even if the three lessons presented are broadly correct, we have to face the more practical or rather pedagogical question whether they should be part of an avalanche education curriculum. Let me finish by highlighting a number of challenges for avalanche educators willing to take the Bayesian perspective seriously.

There is a genuine danger that some students will draw the *wrong* consequences from lessons one and two when the avalanche danger is low or moderate: they might consider stability tests as irrelevant in that context. So they might think that a test indicating a not-safe slope is not informative enough and so they might be tempted to ignore the test results.

This kind of thinking will have to be countered effectively and to do so educators will likely need to present the lessons in the wider context of approaching avalanche education as an efficient way to reduce the risk more generally. As alluded to above, while a 2% probability of triggering an avalanche (i.e. skiing a slope that tested not-safe on a moderate day) might not seem high to some, it will likely seem much more unacceptable if that risk can easily be reduced to 0.3% by skiing a different slope. To strengthen this observation, educators may also want to highlight the cumulative aspect of risk: taking higher risks on a regular basis will make it much more likely in the long run to get caught in an avalanche than always choosing the lower risk (Ebert & Photopoulou, 2013).

Alternatively, to further foreclose the above misinterpretation, educators may heed the lessons from our discussion in Section 4.3 and remind their students not to read too much into the specific numbers, since personal and geographical circumstances are often ignored in such general considerations. Additionally, in the context of regarding stability tests as tools for risk reduction, it will be important for educators to highlight that different kinds of tests at different locations will help increase the hit rate and so further reduce the chances of triggering an avalanche.

Lastly, educators may decide to introduce the Bayesian perspective and its three lessons in more advanced avalanche education courses only - courses that are usually taken by aspiring avalanche forecasters or mountain guides. Given that most introductory courses to avalanche decision-making focus more on avalanche awareness rather than the use of various stability tests, Bayesian reasoning will become much more relevant at a more advanced level of avalanche decision-making. When doing so, it will be prudent to take note of Gigerenzer and Hoffrage (1995), who showed how different information formats and presentations of the problem can help to induce better Bayesian reasoning even without detailed training.

Ultimately, avalanche educators have to decide whether the theoretical benefits of the Bayesian approach will outweigh the potential disadvantages of misinterpreting them. Needless to say, it is beyond this discussion to make that decision for avalanche educators. However, given Eddy's (Eddy, 1982) call to train medical experts of the potential pitfalls of ignoring base-rate information so to avoid 'major errors' when advising their patients, maybe mountain guides and advanced avalanche decision-makers could benefit from learning more about similar pitfalls that potentially affect their decision-making in avalanche terrain.

Notes

1. Moreover, subjects are easily confused about *conditional* probability: instead of considering the probability that a women has breast cancer *given* a positive test result (8%), subjects often tend to consider the probability that you have a positive test result *given* that you have breast cancer (80%) or at least a probability somewhat close to it (Bar-Hillel, 1980; Eddy, 1982). In the legal literature on the Base rate fallacy a similar confusion about conditional probabilities is referred to as the *prosecutor's fallacy* (Thompson & Shumann 1987). There are now a number of theories as to the mechanism that underwrite wrong answers. See in particular Pennycook and Thompson (2017) for a survey of the different theories.
2. Compare Bar-Hillel (1980, p. 215) 'the genuineness, the robustness, and the generality of the base-rate fallacy are matters of established fact'.
3. This is broadly in line with, for example (Munter, 2003) 3×3 method, and the now widely adopted approach that encourages backcountry skiers to start with general information and to update continuously throughout the day on more and more specific information relating to the indented slope that is to be skied (Tremper, 2008).
4. This consideration is very much inspired by Eddy (1982), who makes a very similar point in the context of medical diagnostic tests.
5. The notion of a 'wicked' learning environment is due to Hogarth (2001). The general issue of drawing the wrong conclusion from misleading feedback in avalanche terrain is also highlighted in Zweifel and Haegeli (2014) and Ebert (2015).
6. Compare here Munter (2003), who introduced the idea of avalanche decision-making as a risk reduction method. As such this lessons is not new but the way we arrive at it, by considering Base rate information and applying Bayes' theorem, offers a new way to strengthen this important observation.
7. Note that strictly speaking the risk is more than 100 micromort since we are only looking at fatalities from avalanches and not accidents more generally.
8. Further discussion can be found in Tremper (2013), who suggests that not using any risk reduction measures on a moderate avalanche level incurs roughly 20 micromort (assuming 10 slope crossings) and on a considerable warning roughly 300 micromort (assuming 10 slope crossing), while risk reduction methods can reduce it to 2–4 micromort (when ski touring in Austria) irrespective of the warning. So compared with this informal survey our estimates make ski touring much more dangerous than usually thought. Also note that Winkler (2016) suggests that the risk of skitouring per day is roughly 9 micromort in Switzerland. For difficulties interpreting general risk assessments involving mountain sports, see also (Ebert & Roberson 2013) and the discussion below.
9. It's worth highlighting that if we lower the risk factor (currently 4) that is associated with a single increase of the general avalanche warning (say from low to moderate), then this might reduce the effect alluded to in lesson 1 but it does not cancel it out so long as the factor is greater than 1.
10. Also, remember that a test might be perfectly accurate yet we might still make mistakes interpreting the test, e.g. we might chose the wrong location for a test, subjects can be affected by confirmation bias, etc.
11. There is always the danger that *perceived* competence in reducing the risks, will lead to an increase in a subject's willingness to take risk. Llewellyn et al. (2008) showed this in the case of rock climbing. It is

tentatively suggested for ski-touring in Zweifel (2012). Additional problems arise when the *perceived* competence of a avalanche decision-maker is not grounded in *genuine* competence (Ebert, 2015).

12. This issue is a version of a more general phenomenon called the *reference class problem* in the philosophy of science and statistics literature, compare (Hájek, 2007).
13. Yet one further step would be appeal to second-order probabilities to capture the idea that there is a degree of epistemic uncertainty or risk about our first-order probabilities (see Gärdenfors and Sahlin 1982; Sahlin, 1983). While this might be a very promising and highly relevant approach, especially in the context of behavioural avalanche decision-making, I will leave further discussion to another occasion.

Acknowledgement

I would like to thank Karl Birkeland, Ian Durbach, Nora Hanson, Peter Milne, Matthias Sammer, Martin Smith, Frank Techel, Bruce Tremper and Benjamin Zweifel for their critical and immensely helpful comments on earlier versions of this article. In addition, I would like to thank the audience of a workshop at the *Center for Avalanche Research and Education*, University of Tromsø, for their valuable feedback. Special thanks to Audun Hetland, Andrea Mannberg, and Michael Morreau for organising the event. Lastly, I would like to thank the anonymous referees for their time and their helpful and encouraging comments.

Disclosure statement

No potential conflict of interest was reported by the author.

Notes on contributor

Philip Ebert is as a Senior Lecturer in Philosophy at the University of Stirling, Scotland. One of his specialisation is in Epistemology with interests in norms of reasoning, (Bayesian) rationality, and risk. He spends his spare time in the mountains: ski-touring, climbing, mountain-biking.

ORCID

Philip A. Ebert http://orcid.org/0000-0002-3730-0600

References

Jamieson, B., Schweizer, J., & Shea, C. (2009). Simple calculations of avalanche risk for back-country skiing. *Proceedings of the 2009 International Snow Science Workshop*, Davos, pp. 336–340.

Barbey, A. K., & Sloan, S. A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioural and Brain Sciences*, *30*, 241–297.

Bar-Hillel, M. (1980). Base-rate fallacy in probability judgements. *Acta Psychologica*, *44*, 211–233.

Birkeland, K. W., & Chabot, D., 2006. Minimizing 'false stable' stability test results: Why digging more snowpits is a good idea. *Proceedings of the 2006 International Snow Science Workshop*, Telluride, Colorado, pp. 498–504.

Birnbaum, M. H. (2004). Bayes rates in Bayesian inference. In R. F. Pohl (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory* (pp. 43–60). New York, NY: Psychology Press.

Brugger, H., Etter, H., Zweifel, B., Mair, P., Hohlrieder, M., Ellerton, J., … Falk, M. (2007). The impact of avalanche rescue devices on survival. *Resuscitation*, *75*(3), 476–483.

Casscells, W., Schoenberger, A., & Grayboys, T. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine*, *299*, 999–1000.

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., & Photopoulou, T. (2013) Bayes' beacon: Avalanche prediction, competence and evidence for competence. Modelling the effect of competent and incompetent predictions of highly improbable events. *Proceedings of the 2013 International Snow Science Workshop*, Grenoble, France, pp. 363–370.

Ebert, P. A., & Robertson, S. (2013). A plea for risk. *Royal Institute of Philosophy Supplement*, *73*, 45–64.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine. Problem and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 249–267). Cambridge: Cambridge University Press.

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.

Gärdenfors, P., & Sahlin, N.-E. (1982). Unreliable probabilities, risk taking, and decision making. *Synthese*, *53*, 361–386.

Gauthier, D., & Jamieson, J. B. (2008). Fracture propagation propensity in relation to snow slab avalanche release: Validating the propagation saw test. *Geophysical Research Letters*, *35*(13). doi:10.1029/2008GL034245

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684–704.

Grímsdóttir, H., & McClung, D. (2006). Avalanche risk during backcountry skiing. An analysis of risk factors. *Natural Hazards*, *39*, 127–153.

Hájek, A. (2007). The reference class problem is your problem too. *Synthese*, *156*(3), 563–585.

Hendrikx, J., & Birkeland, K. W. (2008). Slope scale spatial variability across time and space: Comparison of results from continental and maritime climates. *Proceedings of the 2008 International Snow Science Workshop*, Whistler, B.C., pp. 155–162.

Hendrikx, J., Birkeland, K. W., & Clarke, M. (2009). Assessing changes in the spatial variability of the snowpack fracture propagation propensity over time. *Cold Regions Science and Technology*, *56*(2–3), 152–160.

Hogarth, R. M. (2001). *Educating Intuition*. Chicago: University of Chicago Press.

Howard, R. A. (1980). On making life and death decisions. In *Societal risk assessment: How safe is safe enough?* J. Richard, C. Schwing, & W. A. Albers (Eds.), General Motors Research Laboratories. New York: Plenum Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237–251.

Kipps, C., Sharma, S., & Pedoe, D. T. (2011). The incidence of exercise-associated hyponatraemia in the London marathon. *British Journal of Sports Medicine*, *45*(1), 14–19.

Leiter, A. M. (2011). The sense of snow – Individuals' perception of fatal avalanche events. *Journal of Environmental Psychology*, *31*, 361–372.

Llewellyn, D. J., Sanchez, X., Asghar, A., & Jones, G. (2008). Self-efficacy, risk taking and performance in rock climbing. *Personality and Individual Differences*, *45*(1), 75–81.

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*(April), 78–85.

McCammon, I. (2004). Heuristic traps in recreational avalanche accidents: Evidence and implications. *Avalanche News*, *68*, 42–50.

McClung, D. M. (2011). The strength and weight of evidence in backcountry avalanche forecasting. *Natural Hazards*, *59*, 1635–1645.

McClung, D. M., & Schaerer, P. (1993). *The avalanche handbook*. Seattle: Mountaineers Books, USA.

Munter, G. (2003). *3 × 3 Lawinen. Risikomanagement im wintersport* (3rd ed.). Garmisch-Partenkirchen, Germany: Pohl und Stellhammer.

Pennycook, G., & Thompson, V. A. (2017). Base rate neglect. In R. F. Pohl (Ed.), *Cognitive illusions. Intriguing phenomena in thinking, judgement, and memory* (pp. 44–61). Oxford: Routledge.

Pfeifer, C. (2009). On probabilities of avalanches triggered by alpine skiers. An empirically driven decision strategy for backcountry skiers based on these probabilities. *Natural Hazards*, *48*(3), 425–438.

Sahlin, N.-E. (1983). On second order probabilities and the notion of epistemic risk. In B. P. Stigum & F. Wenstøp (Eds.), *Foundations of utility and risk theory with applications.* (pp. 95–104). Dordrecht: D. Reidel.

Schweizer, J., & Lutschg, M. (2001). Characteristics of human-triggered avalanches. *Cold Regions Science and Technology*, *33*, 147–162.

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*, 210–216.

Soreide, K., Ellingsen, C. L., & Knutson, V. (2007). How dangerous is base jumping? An analysis of adverse events in 20,850 jumps from the Kjerag Massif, Norway. *The Journal of Trauma: Injury, Infection, and Critical Care*, *62*(5), 1113–1117.

Spiegelhalder, D., & Blastland, M. (2014). *The norm chronicles: Stories and numbers about danger*. London, UK: Profile Books.

Techel, F., Zweifel, B., & Winkler, K. (2015). Analysis of avalanche risk factors in backcountry terrain based on usage frequency and accident data in Switzerland. *Natural Hazards and Earth System Science*, *15*, 1985–1997.

Thompson, E. L., & Shumann, E. L. (1987). Interpretation of statistical evidence in criminal trials: The Prosecutor's fallacy and the defense Attorney's fallacy. *Law and Human Behavior*, *11*(3), 167–187.

Tremper, B. (2008). *Staying alive in avalanche terrain* (2nd ed.). Seattle: Mountaineers Books.

Tremper, B. (2013). What is the risk of riding in avalanche terrain?. Blogpost on https://utahavalanchecenter.org/blog-what-risk-riding-avalanche-terrain" (26/2/2013), retrieved on 20/2/2018.
Winkler, K., Fischer, A., & Techel, F. (2016). "Avalanche risk in winter backcountry touring: Status and recent trends" Proceedings of the 2016 International Snow Science Workshop, Breckenridge, Colorado, pp. 270–276.
Zweifel, B. (2012). Jung & wild vs erfahren. *Bergundsteigen*, *4*, 78–82.
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., Räz, A., & Stucki, T. (2006). "Avalanche risk for recreationists in backcountry and in off-piste area: Surveying methods and pilot study at Davos, Switzerland." Proceedings of the 2006 International Snow Science Workshop, Telluride, Colorado, pp. 733–741.

Appendix A. Bayes' theorem

Bayes' theorem is often stated in this form (assuming that $P(D) > 0$):

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

$P(D)$ can be reformulated as such:

$$P(D) = P(D|H)P(H) + P(D|\neg H)(1 - P(H))$$

Giving us the following version of Bayes' theorem:

$$P(H|D) = \frac{P(D|H)P(H)}{P(D|H)P(H) + P(D|\neg H)(1 - P(H))}$$

Now consider again the *avalanche problem*:

> The probability of a skier triggering a decent sized avalanche on any given skiable slope in the area is 1%. If a slope is not-safe, a stability test will have 80% of indicating that it is not-safe (hit rate). If a slope is safe, the probability that the stability test indicates it is not-safe is 10% (false-positive). A stability test is applied to a skiable slope with the result indicating that it is not-safe. What is the probability that the slope is not-safe?

With the following abbreviations:
H= the slope is not-safe.
D= test result indicating not-safe slope.
$P(D|H)$= probability of test result indicating not-safe, given the slope is not-safe.
$P(D|\neg H)$= probability of a test result indicating not-safe, given the slope is safe.

We can translate the above information thus:
$P(H) = 0.01$ (base rate)
$P(D|H) = 0.8$ (hit rate)
$P(D| - H) = 0.1$ (false-positive rate)

And then calculate the result:

$$P(H|D) = \frac{0.8 \times 0.01}{0.8 \times 0.01 + 0.1 \times (1 - 0.01)} = \sim 0.075 = 7.5\%$$