A BAYESIAN PERSPECTIVE ON AVALANCHE DECISION-MAKING AND THE RELEVANCE OF STABILITY TESTS

Philip A. Ebert[1,*]

[1] *Division of Law and Philosophy, University of Stirling, Stirling, UK, and Behavioural Science Centre, University of Stirling, Stirling, UK*

ABSTRACT: In this paper, I explore a Bayesian perspective on avalanche decision-making. I motivate this general outlook by introducing a well-known cognitive bias, the base-rate fallacy, and show how a similar pattern applies to decision-making in avalanche-terrain when assessing the relevance of stability tests. I then present three theoretical lessons that emerge from adopting a Bayesian perspective to avalanche decision-making. I conclude by raising numerous challenges for avalanche educators when incorporating the Bayesian perspective into their curriculum and point to future research.

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 et.al., 2010), (Leiter, 2011), (Marengo et.al, 2017). While much emphasis has been placed on heuristic traps highlighted in (McCammon, 2004), little attention has been paid to a well-known *cognitive* bias concerning reasoning and thinking about probabilities called the *base-rate fallacy*. When committing this fallacy, intuitive thinking involving probabilities is systematically off-target and leads to serious misjudgements about the risk of certain events.

The aim here is not to investigate whether back-country skiers in particular are more or less prone to this particular fallacy, the article is rather more *theoretical* in nature: highlighting the base-rate fallacy and its relevance, I first motivate a broadly Bayesian perspective on avalanche decision-making, in particular in relation to the use of stability tests. In the subsequent section, I highlight three lessons for avalanche decision-making that emerge from adopting the Bayesian perspective. I conclude by highlighting various challenges for avalanche educators when incorporating Bayesian reasoning into their curriculum.

*Corresponding author address:
Philip A. Ebert, Division of Law and Philosophy, University of Stirling, Stirling, FK9 4LA, UK
email: p.a.ebert@stir.ac.uk

2. BAYESIAN REASONING AND THE BASE-RATE FALLACY

The base rate fallacy became prominent through the work of (Kahneman and Tversky, 1973). It occurs when a subject does not take 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 forty 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, 1982)'s 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%-80%. Since then many more formal studies have been conducted using different contexts and the results are broadly speaking similar (for variations, however, see (Gigerenzer and Hoffrage, 1995)). The correct answer in the above case is roughly 8% and usually lower than intuitive estimates.

This form of Bayesian reasoning can, however, be made relevant in an avalanche decision-making context, provided we can identify the base rate for an avalanche event and the relevant hit and false-positive rate of slope stability tests. For simplicity, let's say that a slope is *not-safe* if and only if a skier will trigger a decent sized avalanche on that slope. The required *base rate*, therefore, will be the probability that a slope is not-safe *prior to having applied slope specific diagnostic tests*. This type of data is difficult to acquire without knowing more about the number of people engaged in the activity, the time spent in avalanche terrain, and the competencies of the skiers involved (Techel et.al., 2015), (Winkler et.al., 2016).

However, enough relevant information is available to get a ballpark figure. According to (Winkler et.al., 2016) the fatality rate is around 1:100,000 for skitouring in Switzerland. Using a fatality rate of (very) roughly 1 in 10 people who are caught in an avalanche of decent size (Brugger et.al., 2007) we arrive at a rate of triggering an avalanche of around 1:10,000 (0.01%) per day per skier in that area. Given various "unkowns" in this estimate, let's be safe and use a much higher rate of 1 in 100 (1%) chance that on a given run an avalanche will be triggered *prior* to applying any stability test. I come back to variations of the base rate in section 4.

In order to apply Bayesian reasoning, 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 and most tests currently in use are subject to much interpretation and errors can be due to human fallibility or non-representative sampling (Hendrikx et.al., 2008). Studies suggest so-called "false stable" rates of 6%-44%, which translates to a hit rate of 56% to 94%, and false-positive rates between 0%-18% (Simenhois and Birkeland, 2009). Let's again simplify and use a 80% hit rate and a 10% false-positive rate.

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 a 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?

Just as in the *Mammography Problem*, 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 a skier trigger an avalanche on such a slope, given, of course, the above guesstimates for base rate and accuracy of test. In what follows, we will discuss how this *thought experiment*, which exemplifies Bayesian reasoning, is of value to avalanche decision-makers.

3. THREE LESSONS OF BAYESIANISM

To repeat, my intention here is not to explicitly test whether those navigating through avalanche terrain are more or less susceptible to this kind of base rate neglect. Rather, the above considerations are used to motivate a Bayesian approach to avalanche decision-making. In what follows, I outline three general lessons that emerge from adopting a Bayesian perspective.

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. 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 2-10. (Pfeifer, 2009), (Munter, 2003), (Techel et.al., 2015), (Jamieson et.al., 2009).

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.

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*, or more simply, its *informativeness*, 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 or the informativeness 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:

Lesson 1: the role of base rates
> Localised diagnostic tests will be more *informative* the higher the general avalanche warning.

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:

Lesson 2: 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.

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 chance into a wider context: after all, what if a different slope actually does test safe? Assuming for simplicity, the same base rate and the same hit/false-positive rate as above, the probability that a slope is safe, given the test predicts it is safe, is 99.97% (3 in 1000) which is much less when compared to 98% (20 in 1000). Hence we arrive at our last insight which is line with with (Munter, 2003) idea of risk-reduction.

Lesson 3: the method of risk reduction
> 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.

4. DISCUSSION

Numerous challenges to this approach are discussed in detail in Ebert (2018). Here it is worthwhile highlighting that the Bayesian approach remains relevant even if the adopted base rate is much lower (and so more in line with (Winkler et.al., 2016)), or if the accuracy of the test is much improved. In (Figure 1, Table 1), we look at the effects of raising the hit rate to 95% and lower the false-positive rate to 5% which is better than most tests available today Simenhois and Birkeland (2009).
Considering the three lessons above in the context of these results, we can see that the first lesson is still supported: the higher the avalanche rating the

greater the *informativeness* of the test. Moreover, we can see that an increase in the accuracy of the test, results in an increase of informativeness (all else equal).

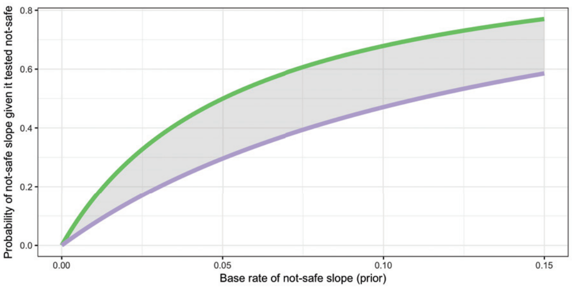

Figure 1: The effects of increasing accuracy of tests on the posterior risk. Purple line represents lower accuracy (80% hit rate; 10% false positive), while green line represents higher accuracy (95% hit rate; 5% false positive). Grey area represents variability of posterior given variation in accuracy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

Avalanche rating	[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%)

Table 1: The effect of increasing accuracy of tests on the posterior risks: the left column contains the base rate. The other two columns present the probability of an avalanche after 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).

The second lesson 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 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 (compared to 1.9% and 4.5% respectively, using the less accurate test). Hence, there is still a danger to misinterpret the most likely outcome (i.e. no avalanche) as evidence that the test is unreliable, which is what lesson two warns against. Note that, if, as alluded to above, the base rate is in fact much lower, then *the problem of misleading feedback* will be even more pronounced.

Lesson three also remains intact and the variation of the accuracy of the test nicely highlights how effective avalanche training can help to reduce the risk–especially if we assume that the better trained the individual, the more accurate her stability tests. Only if our test is maximally accurate will there be no residual risk–a scenario which is extremely (to say the least) unlikely.

5. BAYESIAN REASONING AND AVALANCHE EDUCATION

The theoretical exercise of using Bayesian reasoning contains a number of lessons about how to reason properly with *evidence* provided by stability tests that are not widely acknowledged in avalanche education–the exception being the more technical discussion in (McClung, 2011). Given that Bayesian reasoning is 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 perspective seriously in this context.

Nonetheless, the considerations also contain a huge challenge to avalanche educators: Bayesian reasoning is not always very intuitive (hence the fallacy alluded to above) and it is easily misunderstood. So even if the three lessons presented are correct, we also have to face the more practical or rather pedagogical question whether they should be part of an avalanche education curriculum? I will highlight some 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 lesson 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 skiers), it will likely seem unacceptable if that risk can easily be reduced to 0.3% by skiing a different slope. To strengthen this observation, educator 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 and Photopoulou , 2013).

Alternatively, to further foreclose the above misinterpretation, educators ought to remind their students not to rely too heavily on specific numbers: there is a great variability in the data due to personal, geographic and other factors. Importantly, the greater the variability of the risk, the lower the relevance of the mean average–used in most data above–when considering an individual's risks. Nonetheless, while we need to be very careful when applying the above risk statements to a specific case, there is value in considering the above scenarios as a theoretical exercise since the three lessons of the Bayesian approach will not be affected by this variability. 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 to increase the hit rate (though may increase the false-positive rate) and so can help to further reduce the chances of triggering an avalanche. Further research of how best to *aggregate judgements*, i.e. combine the results from a number of different stability tests or from different subjects, is planned.

Lastly, educators may decide to introduce the Bayesian perspective and its three lessons in more advanced avalanche education courses only. These courses 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 outweigh the potential disadvantages of misinterpreting them. 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 can affect their decision-making in avalanche terrain.

ACKNOWLEDGEMENT

This contribution is a revised and shortened version of Ebert (2018). As in the original article, I wish to thank Karl Birkeland, Ian Durbach, Nora Hanson, Audun Hetland, Andrea Mannberg, Peter Milne, Michael Morreau, Matthias Sammer, Martin Smith, Frank Techel, Bruce Tremper, and Benjamin Zweifel for their critical and helpful comments on earlier versions of this article. In addition, I'd like to thank the audience of a workshop at the *Center for Avalanche Research and Education*, University of Tromsø, for their valuable feedback.

REFERENCES

Brugger, H., Etter, H., Zweifel, B., Mair, P., M., H., Ellerton, J., Elsensohn, F., Boyd, J., Sumann, G., and Falk, M. 2007. The impact of avalanche rescue devices on survival. *Resuscitation*, 75(3), 476-483.

Ebert, P.A. 2018. Bayesian reasoning in avalanche terrain: a theoretical investigation. *Journal of Adventure Education and Outdoor Learning*, https://doi.org/10.1080/14729679.2018.1508356

Ebert, P.A. and 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, 363-370.

Eddy, D.M. 1982 Probabilistic reasoning in clinical medicine. Problem and opportunities. In *Judgement under uncertainty: Heuristics and biases*, edited by Daniel Kahneman, Paul Slovic, Amos Tversky, 249-67. Cambridge: Cambridge University Press.

Furman, N., Shooter, W., and 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.

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

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

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

Kahneman, D., Tversky, A. 1973. On the psychology of prediction. *Psychological Review*, 80, 237-251.

Marengo, D., Monaci, M.G.. and 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. 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.

Munter, G. 2003. *3x3 Lawinen. Risikomanagement im Wintersport*. Pohl und Stellhammer: Garmisch-Partenkirchen. 3rd edition.

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

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.

Simenhois, R. and 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

Techel, F., Zweifel, B., and 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.

Winkler, K., Fischer, A, and Techel, F. 2016. Avalanche risk in winter backcountry touring: status and recent trends. *Proceedings of the 2016 International Snow Science Workshop, Breckenridge, Colorado*, 270-276.