Encouraging Metacognition with Bayesian Grading

Paul Mayer
Dept. of Electrical and Computer Engineering
Rice University
Houston, Texas USA
0009-0007-2801-6601

Lê Nguyên Hoang?

Bailleux Olivier?

Rich Baraniuk
Dept. of Electrical and Computer Engineering
Rice University
Houston, Texas USA
0000-0002-0721-8999

***Abstract*—The vast majority of exam grading, especially in true-false and multiple choice scenarios, encourages students to guess. This leads to two issues: the first is it incentivizes students to be overconfident rather than honestly quantify their level of belief and the second is that instructors cannot see which questions students are certain about versus these where students guess. This paper explores how Bayesian Grading solves these issues, encouraging students to be honest about their true beliefs and discouraging them from guessing. Bayesian Grading evaluates both cognition and metacognitive reflection simultaneously, teaching students to become aware of the sources of their uncertainty and encouraging them to be self-directed learners.**

***Index Terms*—Scoring, Metacognition, Grading, Bayesian Probability**

I. Background

Metacognition is perhaps best summarized as "thinking about thinking," consisting of two components: knowledge and regulation. Metacognitive reflection is particularly relevant when teaching students to become independent thinkers and learners, as it involves "the monitoring of one's cognition and includes planning activities, awareness of comprehension and task performance, and evaluation of the efficacy of monitoring processes and strategies (Lai, 2011)." Research suggests metacognition can also improve with appropriate instruction, implying students can be taught to reflect on their own thinking (Lai, 2011).

While teaching metacognition is associated with "very positive effect[s] on student outcomes (Perry et al., 2019)," it is often underused in classrooms. This is likely due to the difficulty assessing metacognition and integrating it with existing curricula. This paper focuses on a method of modifying true-false and multiple choice questions to incentivize students to practice metacognition and honestly report their level of certainty in answer selections being true. The method proposed in this paper, Bayesian Grading, requires no change to existing multiple choice or true-false questions, instead relying on students reporting their creedences combined with a scoring rule for these creedences that incentivize honesty. Bayesian Grading can be done automatically via a computer[1], requiring no additional instructor effort once students have been taught the process. The process of teaching students to reporting their creedences gives them practice quantifying their knowledge in line with standards of reason described by many epistemologists (Cox, 1961; Hájek, 2023).

II. Motivation: Issues with Multiple Choice Scoring

Consider the following multiple choice question, taken from the FE (Fundamentals of Engineering) exam:

> A magnetic field has the vector field $\vec{B} = 10y\mathbf{i} + 2y\mathbf{j} - mz\mathbf{k}$ The constant m is most nearly:
> - A: 2
> - B: -2
> - C: 0
> - D: 4

Consider the students Alice, Bob, and Charles who answer the question as follows: Alice is 51% confident the answer is A and 49% confident the answer is B, so she chooses A. Bob is 49% confident the answer is A and 51% confident the answer is B, so he chooses B. Finally, Charles has no idea of what the answer is, so he guesses answer A. In typical multiple choice grading, (detailed in Section III-C(students choose one answer they think is the best. They receive 1 point if they select the correct answer and 0 points if they select the wrong answer.

In the above example, Alice and Charles both receive 1 point, while Bob receives 0 points. This is despite the fact that Alice and Bob's underlying (metacognitive) beliefs about the correct answer are very similar. Furthermore, Charles receives more points than Bob by pure luck, despite the fact that Bob has more knowledge than Charles: he is able to eliminate two *incorrect* answers from the problem.

The standard way of grading multiple choice questions, with its all-or-nothing nature, provides no way of distinguishing

Identify applicable funding agency here. If none, delete this.

[1]Code is available at: REMOVED FOR REVIEW PROCESS

between Alice and Charles' knowledge, since both receive the same number of points. An instructor thus has no way of knowing that Charles was guessing and Alice knew the correct answer, and thus no recourse for intervening on Charles' behalf and fixing whatever caused him to believe B was the correct answer. An instructor also has no way of knowing whether Bob was simply guessing or had a misconception that caused him to choose an incorrect answer. The issue with standard multiple choice grading can be summarized as follows: students are not incentivized to report their true beliefs because they have no way of doing so. This harms both students who could display some partial knowledge (for instance, being able to eliminate some incorrect answers) and instructors who want to know which concepts that students are confused about. While standard multiple choice grading does show the questions students get wrong, it does not inform instructors which answers students are deciding between, and thus which distractors students are confused by. The ability to see students' confidence in various answer choices allow instructors to better understand students' misconceptions and thought processes, allowing them to teach more effectively.

A proper way of scoring multiple choice questions should thus have the following desirable properties:

- Offer maximum credit if a student is confident in the correct answer
- Provide partial credit if students can eliminate *some* incorrect answers
- Offer minimum credit if a student is overly confident in an *incorrect* answer (discouraging students from guessing randomly)

In summary, a metacognitively-aimed way of grading multiple choice questions will incentivize students to report their *true* level of confidence in the correct answer, rewarding them the more their confidence aligns with the correct answer, *and* the more confident they are that incorrect answers are incorrect.

III. From Boole to Bayes

A. Quantifying Belief

Using numbers to quantify one's confidence in a statement being true corresponds to the Bayesian interpretation of probability (Hájek, 2023), an extension of Aristotle's logic (Jaynes, 2003; Cox, 1961, 1946). According to Plato and Aristotle (Hamilton, 2009) statements or propositions are either true or false, never both or neither, and with no in-between. These principles about statements are captured in the "three laws of thought," which can be expressed as follows:

1) Law of Non-Contradiction: $\neg(A \wedge \neg A)$
2) Law of the Excluded Middle: $A \vee \neg A$
3) Law of Identity: $\forall x, x = x$

Logician George Boole famously used algebra to describe these axioms (Boole, 1911) and Aristotle's syllogisms more generally (Corcoran, 2003), allowing mathematical and symbolic operations to be performed on logical statements (Davis, 2000). In Boole's formulation, a value of 1 represents a true statement, while a value of 0 represents a false statement.

While Boole's logic is able to capture the truth or falsity of statements, it does not provide a mechanism for evaluating one's knowledge confidence in whether a statement is true or false. This is because Boolean logic is aimed at describing truth, not knowledge. Bayesian probability is thus an extension of logic that captures an agent's subjective degree of belief in a statement's truth (Hájek, 2023). Furthermore, as physicist Richard Cox famously pointed out, if one takes (as Boole did) 0 to represent false and 1 to represent true, the probability measure P in a Kolmogorov probability triple (Ω, E, P) (Kolmogorov and Bharucha-Reid, 2018) that assigns numeric values to events $e \in E$ corresponds to Boole's truth values when one has complete certainty in an event occurring $P(e) = 1$ or complete certainty in an event not occurring $P(e) = 0$ (Cox, 1961, 1946). Because the probability measure can take on intermediate values between 0 and 1, it can capture situations where an agent lacks knowledge about which outcome occurs (corresponding to a scenario where an agent lacks knowledge about whether the proposition "an event e_i occurs" is true). Interpretations of $P(e_i) = p$ are thus "I am $p \cdot 100\%$ confident event e_i occurs" or "if I had x dollars to bet on e_i occurring, I would bet $p \cdot x$ dollars (Ramsey, 2016)."

Such values do not contradict any of the laws of thought discussed in Section III-A because Bayesian probability is an *epistemic* notion, while (logical) truth is an *ontic* one. In other words, one's degree of belief is based on one's current knowledge (and ignorance), while logical truth is a mind-independent property of reality. It is for this reason many epistemologists see (Bayesian) probability as an addition to logic, not a replacement for it.

B. Boolean vs. Bayesian True/False Questions

Questions where students are asked to report whether a statement is true or false are traditionally scored using Boole's logic. Consider the following question:

> True or False: When integrating a function $f_1(x) + f_2(x)$, the integral of the sum $\int (f_1(x) + f_2(x))\,dx$ equals the sum of the integrals $\int f_1(x)dx + \int f_2(x)dx$.

Let k be a boolean variable representing the answer key, meaning it is 0 if the statement is false or 1 if the statement is true. In traditional (what the authors call Boolean) grading, a student reports whether they believe the statement is true or false. Let a represent a students' answer: a student will submit $a = 0$ if they believe the statement is false and $a = 1$ if they believe the statement is true. The score the student receives for a true-false question is $\text{Score}_{\text{BooleanTF}} = k \iff a$, which takes on a value of 1 if k and a have the same value and 0 if they differ. This can also be written as[2]:

$$\text{Score}_{\text{BooleanTF}}(a) = \begin{cases} a & \text{if } k = 1 \\ (1-a) & \text{if } k = 0 \end{cases} \quad (1)$$

[2]Equation 1 can be written succinctly as $\text{Score}_{\text{BooleanTF}}(a) = ak + (1 - a)(1 - k)$.

In principle, this grading scheme incentivizes students to choose whichever answer they believe is best, as the score quantifies the alignment of their answer choice a with the truth k. The issue, however, is that students cannot fully represent their beliefs. Consider two students, Jacob and Karen. Jacob is 95% sure the statement is true, while Karen is only 60% sure the statement is true. Since both must report true or false (representing full confidence in whether they believe the statement is true or false) they have no way of qualifying their belief to take into account their uncertainty. As a result, students are not incentivized to exercise metacognitive regulation, since they have no choice but to report full confidence in a single answer. Furthermore, since an instructor only sees whether a student reports true or false for a given question, they have no way of knowing whether students are simply guessing or are confident that they have the correct answer.

Bayesian Grading, on the other hand, scores questions based on students' creedence c. In addition to capturing cases when a student is fully confident that a given statement is true ($c = 1$) or false ($c = 0$), Bayesian Grading captures when students have no idea whether the it is true or false ($c = 0.5$) along with various positions in-between.

C. Boolean vs. Bayesian Multiple Choice

Typical multiple choice grading also implicitly follows Boole's system of quantifying true and false into 0 and 1. A multiple choice question with n selections is a collection of n statements: one of which is true, and the remaining $n - 1$ statements are false. Students are tasked with selecting the statement they believe is true (and thus which remaining statements, which remain unselected, are false).

Since there is only one statement which is true, the answer key for a multiple choice question can be thought of as a basis vector e_i, a vector of a single 1 and several 0's, where the position of the 1 corresponds to the correct (or true) answer and the positions of the 0's refer to incorrect (or false) choices. For instance, the answer key for the example question in Section II can be written as $\vec{k} = [1, 0, 0, 0]^T$, with the first position corresponding to answer choice A, the second position corresponding to answer choice B, etc. If t refers to the position of the true (correct) answer, then the answer key $\vec{k} = e_t$, a basis vector with a 1 in the t position, and 0's in all other positions.

A students' answer choice can also be thought of as a vector; one that encodes the students belief in what the best answer is. In standard or Boolean multiple choice grading, students must select one answer they believe is correct. As a result, their choice must follow the same rules as the answer key: a 1 in the position of the answer they believe is correct, and a 0 in all other positions. Alice and Charles' answer choices for the example question in Section II can be written as $\vec{a}_{\text{Alice}} = \vec{a}_{\text{Charles}} = [1, 0, 0, 0]^T$ while Bob's answer can be written as $\vec{a}_{\text{Bob}} = [0, 1, 0, 0]$.

A students' score for a given question is thus the inner product of the student's answer $\vec{a}$ with the answer key $\vec{k}$:

$$\text{Score}_{\text{BooleanMC}}\,(\vec{a}) = \langle \vec{a}, \vec{k} \rangle = \sum_i \vec{a}_i \vec{k}_i. \tag{2}$$

This inner-product grading scheme allocates points to students based on how well their beliefs line up with the truth. This approach is unproblematic; instructors want to know what students' knowledge is, and the correspondence between student answers and the truth is a justifiable metric (Glanzberg, 2023). The issue, rather, is students are unable to fully represent their beliefs if they have any uncertainty about the correct answer.

D. Disbelief vs Unbelief

As sociologist of science Marcello Truzzi pointed out, there is a difference between *disbelief* and *unbelief* (Truzzi, 1987). The idea that one's "default position" should be unbelief (reserving judgment about facts) rather than disbelief (assuming a statement is false prior to investigation) dates back all the way to the Greek skeptic Pyrrho, who wrote that one should "be without beliefs, disinclined to take a stand one way or another (Stough, 1969)." After all, claiming a proposition is false is a claim to knowledge in the same way claiming a proposition is true is a claim to knowledge.

When one is completely ignorant of whether a statement is true or false, a creedence of $c = 0.5$ is assigned to the statement to represent unbelief. This corresponds to what is called the "principle of insufficient reason"[3]. The principle of insufficient reason says that in the absence of any relevant evidence, one should distribute their credence equally among all possibilities as to not unjustifiably favor one possibility over another[4]. The state of unbelief represents an openness to whether a statement is true or false, and likewise $c = 0.5$ is equally between disbelief ($c = 0$) and full confidence ($c = 1$).

For multiple choice questions, the principle of indifference says to assign each answer a creedence of $\frac{1}{n}$ when one is completely ignorant of which answer is correct. Represented as a vector, $\vec{u}_n$, is an n-length vector where each element has a value of $1/n$. Teaching students to reflect on their ignorance is key for them to become self-directed learners.

IV. Bayesian Scoring

Rather than forcing students to take an all-or-nothing approach to answering question, Bayesian Grading uses a scoring rule on students' creedences c. A well-designed scoring rule will encourage both cognition (whether students know the correct answer) and metacognition reflection (how confident students are in their choice) simultaneously. Such a scoring rule would have the following desirable properties:

[3]This is also known as the "principle of indifference" by J.M. Keynes (Keynes, 2013).

[4]Applied naively, the principle of indifference can lead to a series of "paradoxes" when the reference class has an uncountably infinite number of members (Hájek, 1996; Kendall and Moran, 2012). Since no exam can ever been administered with this many statements, this point is irrelevant to the given context, and rather serves as a warning for an unscrupulous use of the principle of indifference.

- Offer a maximum score to students who reports full confidence in the correct answer.
- Offer a minimum score to students who reports full confidence in an incorrect answer
- Offer some score in-between for students who report that they are unsure of the correct answer (with various values in-between).
- Incentivize students to honestly report their true creedences (not guess or exaggerate)

A. Naive Bayesian Scoring

One straightfoward way of scoring belief (that adheres to the first three properties above) is to use students' creedences directly in Equation 1 and 2 for true-false and multiple choice questions respectively. For true-false questions, the score of a student's reported creedence c for a question where k is a boolean variable representing the correct answer:

$$\text{Score}_{\text{NaiveBayesTF}}(c) = \begin{cases} c & \text{if } k = 1 \;\; (k \text{ is true}) \\ (1-c) & \text{if } k = 0 \;\; (k \text{ is false}) \end{cases} \tag{3}$$

Similarly, for multiple choice questions, students' creedences $\vec{c}$ in each answer can be compared with the answer key $\vec{k}$ as follows:

$$\text{Score}_{\text{NaiveBayesMC}}(\vec{c}) = \langle \vec{c}, \vec{k} \rangle \tag{4}$$

The problem with this method is that students can maximize their score by being *dishonest* about their true creedences. The fourth desired proprerty of Bayesian scoring is that students are incentivized to report their true creedences. However the expected value of Equations 3 and 4 can be maximized by students who report an exaggerated belief in whichever answer they think is best.

B. Naive Bayesian Scoring for True-False Questions

For a true-false question, the expected score for a reporting strategy is defined as follows, where s is the number of points one allocates and c is one's creedence that the statement is true (i.e. one's beief that the answer key k is true):

$$\mathbb{E}[\text{Score}_{\text{NaiveBayesTF}}(s)] = s \cdot c + (1-s) \cdot (1-c) \tag{5}$$

Consider the question in Section III-B, where the answer key is $k = 1$ (i.e. k is true). Karen is only 60% sure the answer is true, and if she reports her creedence $c = 0.6$, her expected score is as follows:

$$\mathbb{E}[\text{Score}_{\text{NaiveBayesTF}}(0.6)] = 0.6 \cdot 0.6 + 0.4 \cdot 0.4 = 0.52$$

However if Karen instead reports an exaggerated guess using the strategy

$$s_{\text{Exaggerated}} = \begin{cases} 1 & \text{if } c > 0.5 \\ 0 & \text{else} \end{cases}$$

her expected score is

$$\mathbb{E}[\text{Score}_{\text{NaiveBayesTF}}(s_{\text{Exaggerated}})] = 0.6 \cdot 1 + 0.4 \cdot 0 = 0.6$$

Since $\mathbb{E}\left[\text{Score}_{\text{NaiveBayesTF}}(s_{\text{Exaggerated}})\right] > \mathbb{E}\left[\text{Score}_{\text{NaiveBayesTF}}(c)\right]$, the naive scoring method $\text{Score}_{\text{NaiveBayesTF}}$ does not incentivize Karen to honestly report her true level of belief.

C. Naive Bayesian Scoring for Multiple Choice Questions

Similarly, the multiple choice example, consider a student Dan whose creedences for the question described in Section II are: $\vec{c}_{\text{Dan}} = [0.4, 0.3, 0.2, 0.1]$. If he reports his true creedences, his score will be $\langle \vec{c}_{\text{Dan}}, \vec{k} \rangle = 0.4 \cdot 1 + 0.3 \cdot 0 + 0.2 \cdot 0 + 0.1 \cdot 0 = 0.51$. However, if Dan instead reports maximum confidence in the answer he believes is correct, represented by the response $\vec{s}_{\text{Dan}} = [1, 0, 0, 0]$, his score will be $\langle \vec{s}_{\text{Dan}}, \vec{k} \rangle = 1 \cdot 1 + 0 \cdot 0 + 0 \cdot 0 + 0 \cdot 0 = 1$.

If one considers Dan's expected score under each method of choosing an answer, we find that Dan's optimal strategy is *not* to report his true level of belief. The expected score under a reporting strategy $\vec{s}$ is defined as follows:

$$\mathbb{E}[\text{Score}_{\text{NaiveBayesMC}}(\vec{s})] = \langle \vec{s}, \vec{c} \rangle \tag{6}$$

First, consider Dan's expected score by honestly reporting his true level of belief:

$$\mathbb{E}[\text{Score}_{\text{Naive}}(\vec{c})] = 0.4 \cdot 0.4 + 0.3 \cdot 0.3 + 0.2 \cdot 0.2 + 0.1 \cdot 0.1 = 0.3$$

Now, consider Dan's expected score if he uses an exaggerated strategy similar to Karen's above, which reports full confidence in whichever answer he believes is more likely than the others to be true :

$$s_{\text{Exaggerated}} = e_m, \quad m = \arg\max_i \vec{c}_i$$

The expected score of the exaggerated strategy is

$$\mathbb{E}[\text{Score}_{\text{NaiveBayesMC}}(s_{\text{Exaggerated}})] = 1 \cdot 0.4 + 0 \cdot 0.3 + 0 \cdot 0.2 + 0 \cdot 0.1 = 0.4$$

Since $\mathbb{E}[\text{Score}_{\text{NaiveBayesMC}}(s_{\text{Exaggerated}})] > \mathbb{E}[\text{Score}_{\text{NaiveBayesMC}}(\vec{c})]$, we see that if Dan wants to maximize $\text{Score}_{\text{NaiveBayesMC}}$, his optimal strategy is to report overconfidence in whichever answer choice he believes is most likely to be true. This does not incentivize Dan to develop healthy metacognitive reflection about his beliefs - in fact it does the opposite, encouraging him to be overconfident in whichever answer he thinks is best and underconfident in other answers.

D. Strictly Proper Scoring

A scoring mechanism where an agent maximizes their score if and only if they are honest about their true level of belief is called a *strictly proper scoring rule* (Staël von Holstein, 1970; Bickel, 2010). Such a scoring rule ensures that only

$$\forall \vec{s} \neq \vec{c}, \quad \mathbb{E}\left[\text{Score}_{\text{StrictlyProper}}(\vec{s})\right] < \mathbb{E}\left[\text{Score}_{\text{StrictlyProper}}(\vec{c})\right]$$

There are many examples of strictly proper scoring rules; for instance, Bickel (2010) suggests using what is called a logarithmic scoring rule, $L(\vec{c}) = \ln(\vec{c}_t)$, which scores questions based on the logarithm of the creedence a student assigns for the true answer t[5]. One potential issue with this metric is if a student assigns a creedence of 0 to the correct

[5]Such a score would need to be scaled by some offset, since the maximum score would be 0

answer (i.e. $\vec{c}_t = 0$), their score would be $-\infty$. While Bickel (2010) argues this is a feature and not a bug, the authors of this paper suggest a different rule that avoids this potential issue[6].

E. Quadratic Scoring

1) Quadratic Scoring for True-False Questions: The naive scoring in Section IV-A can be modified to produce a strictly proper scoring rule by squaring the distance from a student' creedence c to the correct answer k. In essence, the square function discourages exaggerated guesses by penalizing large deviations from the correct value, and produces a proper scoring rule. For true-false questions, the quadratic scoring rule is as follows:

$$\text{Score}_{\text{QuadraticTF}}(c) = 1 - 2(c-k)^2 \tag{7}$$

Consider Karen, who is only 60% confident that the given statement is true. Under the quadratic scoring rule, Karen gets a higher score for being honest and reporting her true creedences, as shown below:

$$\mathbb{E}\left[\text{Score}_{\text{QuadraticTF}}(c)\right] = \left(1-2(0.6-1)^2\right)\cdot 0.6 + \left(1-2(0.6-0)^2\right)\cdot 0.4 = 0.52$$

On the other hand, if Karen attempts to exaggerate her guess by reporting full confidence in her belief that the statement is true, she gets a lower expected score:

$$\mathbb{E}\left[\text{Score}_{\text{QuadraticTF}}(s_{\text{Exaggerated}})\right] = \left(1-2\cdot(0)^2\right)\cdot 0.6 + \left(1-2\cdot(1)^2\right)\cdot 0.4 = 0.20$$

Thus under the quadratic rule, students are incentivized to report their true level of belief for maximum points, encouraging metacognitive reflection.

2) Quadratic Scoring for Multiple Choice Questions: As with true-false questions, Equation 2 can easily be modified to produce a strictly proper scoring rule by squaring the distance from students' creedences to the correct answer. Since there are multiple answer choices, the 2-norm is used to calculate the distance in the quadratic scoring rule as follows:

$$\text{Score}_{\text{QuadraticMC}}(\vec{c}) = 2\vec{c}_t - \|\vec{c}\|_2^2 \tag{8}$$

In the equation above, t is the index of the true answer, and thus $\vec{c}_i$ is the creedence the student assigns to the correct answer. This equation equals $1 - \left\|\vec{c}-\vec{k}\right\|_2^2$, which can be understood as how close the students' creedences $\vec{c}$ are to the answer key $\vec{k}$. Here $d = \left\|\vec{c}-\vec{k}\right\|_2^2$ represents the distance between the student's answer and the correct answer, while $1-d$ represents how "close" of the student's creedence is to the correct answer.

Consider Dan in Section IV-A, whose creedence was $\vec{c}_{\text{Dan}} = [0.4, 0.3, 0.2, 0.1]$. If Dan accurately reports his true creedences, his expected score using Equation 8 is $\mathbb{E}[\text{Score}_{\text{QuadraticMC}}(\vec{c}_{\text{Dan}})] = 0.3$ However, if Dan instead reports full confidence in whichever answer he believes is most likely, his expected score is $\mathbb{E}[\text{Score}_{\text{Quadratic}}(\vec{c}_{\text{Dan}})] = -0.2$. Not only is Dan incentivized to report his true creedences, he is *penalized* for making an exaggerated guess.

[6] While Bickel's justification of the danger of a student getting a $-\infty$, may be justified for graduate students, such a risk may not be appropriate for undergraduate or primary school students.

Choices Eliminated	2 Choices	3 Choices	4 Choices	5 Choices
0	1/2	1/3	1/4	1/5
1	1 (Correct)	1/2	1/3	1/4
2	-	1 (Correct)	1/2	1/3
3	-	-	1 (Correct)	1/2
4	-	-	-	1 (Correct)

TABLE I
STUDENTS' SCORES WHEN ELIMINATING INCORRECT ANSWERS ACCORDING TO $\text{SCORE}_{\text{QUADRATICMC}}(\vec{c})$ DESCRIBED IN EQUATION 8.

Returning to the other students described in Section II, and using each students' creedeces, one has $\vec{c}_{\text{Alice}} = [.51, .49, 0, 0]^T$, leading to an expected score of 0.520 points. Note that if Alice tried to put full creedence in whatever answer she thought was most likely, her expected score would be 0.020. As a result, Alice is incentivized to report her true creedences, as this strategy yields more expected points for her then guessing. Bob's creedences, $\vec{c}_{\text{Bob}} = [.49, .51, 0, 0]^T$ leads to an expected score of 0.480 points. Lastly, Charles' creedences, $\vec{c}_{\text{Charles}} = [0.25, 0.25, 0.25, 0.25]$ leads to an expected score of 0.250. What if Charles had guessed? If he guesses the correct answer, which he does 25% of the time, he receives 1 point. However if we guesses an incorrect answer, which he does 75% of the time, he gets -1 point. Charles' expected score for guessing is actually -2 points, while his expected score for being honest is 0 points. Charles is thus incentivized to report that he does not know the answer than try and guess randomly.

V. Teaching Metacognition Through Bayesian Grading

With Bayesian grading, students are actively discouraged from black-and-white thinking, which has been associated with emotional resistance to intellectual development (Ambrose et al., 2010). The introduction of Bayesian grading in a classroom provides an opportunity for students to practice metacognition, as well as learn topics at the intersection of probability, epistemology, and cognitive psychology. Part of the challenge is simply students' lack of exposure to metacognition: as psychologist Hartman (2001) noted, "many students are unaware of the concept of metacognition and do not reflect on their thinking and learning strategies and attitudes and how they might be improved." The benefit of Bayesian Grading is that students become familiar with (and practice) assessing their own knowledge and ignorance. While many existing grading schemes penalize students most severely for admitting ignorance (actively discouraging healthy metacognitive habits), Bayesian Grading's incentive is towards accurate reflection.

One well-documented metacognitive challenge is the Dunning-Kruger effect, where students who lack knowledge are overconfident in wrong answers (Dunning, 2011). Research has shown metacognitive differences can contribute to the Dunning-Kruger effect (McIntosh et al., 2019), implying teaching students metacognitive practices can reduce their succeptibility to the Dunning-Kruger effect. Other research suggests intellectual humility, or "the degree to which people recognize that their beliefs might be wrong" also reduces one's susceptibility to the Dunning-Kruger effect (Leman et al., 2023). This suggests that Bayesian Grading, which has students report their creedences and provides feedback on these creedences, could reduce the Dunning-Kruger effect in students.

VI. Conclusion

With Bayesian grading, students are incentivized to learn the material *and* learn to quantify their own belief. Bayesian grading thus evaluates cognition and metacognitive reflection simultaneously, with students' creedences allowing instructors to see which material is causing student confusion.

Acknowledgment

TBD

References

E. R. Lai, "Metacognition: A literature review," *Always learning: Pearson research report*, vol. 24, pp. 1–40, Apr. 2011.

J. Perry, D. Lundie, and G. Golder, "Metacognition in schools: what does the literature suggest about the effectiveness of teaching metacognition in schools?" *Educational Review*, vol. 71, no. 4, pp. 483–500, Jul. 2019, publisher: Routledge _eprint: https://doi.org/10.1080/00131911.2018.1441127. [Online]. Available: https://doi.org/10.1080/00131911.2018.1441127

R. T. Cox, *The Algebra of Probable Inference*. Baltimore: Johns Hopkins Press, 1961.

A. Hájek, "Interpretations of Probability," in *The Stanford Encyclopedia of Philosophy*, winter 2023 ed., E. N. Zalta and U. Nodelman, Eds. Metaphysics Research Lab, Stanford University, 2023. [Online]. Available: https://plato.stanford.edu/archives/win2023/entries/probability-interpret/

E. T. Jaynes, *Probability Theory: The Logic of Science*. Cambridge University Press, Apr. 2003, google-Books-ID: tTN4HuUNXjgC.

R. T. Cox, "Probability, Frequency and Reasonable Expectation," *American Journal of Physics*, vol. 14, no. 1, pp. 1–13, Jan. 1946, 781 citations (Crossref/DOI) [2024-11-15]. [Online]. Available: https://doi.org/10.1119/1.1990764

W. Hamilton, *Lectures on Metaphysics and Logic: Vol 1*, Apr. 2009. [Online]. Available: https://www.amazon.com/Lectures-Metaphysics-Logic-V-1/dp/B002NGNBC8

G. Boole, *The Laws of Thought (1854)*. Open Court Publishing Company, 1911.

J. Corcoran, "Aristotle's Prior Analytics and Boole's Laws of Thought," *History and Philosophy of Logic*, vol. 24, no. 4, pp. 261–288, Dec. 2003, publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01445340310001604707. [Online]. Available: https://doi.org/10.1080/01445340310001604707

M. Davis, *The Universal Computer: The Road from Leibniz to Turing*, first edition ed. New York, NY London: W. W. Norton & Company, Oct. 2000.

A. N. Kolmogorov and A. T. Bharucha-Reid, *Foundations of the Theory of Probability: Second English Edition*. Courier Dover Publications, Apr. 2018.

F. P. Ramsey, "Truth and Probability," in *Readings in Formal Epistemology: Sourcebook*, H. Arló-Costa, V. F. Hendricks, and J. van Benthem, Eds. Cham: Springer International Publishing, 2016, pp. 21–45. [Online]. Available: https://doi.org/10.1007/978-3-319-20451-2_3

M. Glanzberg, "Truth," in *The Stanford Encyclopedia of Philosophy*, fall 2023 ed., E. N. Zalta and U. Nodelman, Eds. Metaphysics Research Lab, Stanford University, 2023. [Online]. Available: https://plato.stanford.edu/archives/fall2023/entriesruth/

M. Truzzi, "Zetetic Ruminatrons on Skepticism and Anomalies in Science," *Citeseer*, vol. 12-13, pp. 7–21, 1987.

C. L. Stough, *Greek Skepticism; a Study in Epistemology*. Berkeley,: University of California Press, 1969.

J. M. Keynes, *A Treatise on Probability*. Courier Corporation, Sep. 2013, google-Books-ID: rVCoAAAAQBAJ.

A. Hájek, ""Mises redux" — Redux: Fifteen arguments against finite frequentism," *Erkenntnis*, vol. 45, no. 2, pp. 209–227, Nov. 1996, 57 citations (Crossref/DOI) [2025-01-18]. [Online]. Available: https://doi.org/10.1007/BF00276791

M. G. Kendall and P. A. P. Moran, *Geometrical Probability*. Ulan Press, Sep. 2012.

C.-A. S. Staël von Holstein, "Measurement of Subjective Probability," *Acta Psychologica*, vol. 34, pp. 146–159, Jan. 1970. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0001691870900132

J. E. Bickel, "Scoring Rules and Decision Analysis Education," *Decision Analysis*, vol. 7, no. 4, pp. 346–357, Dec. 2010, 23 citations (Crossref/DOI) [2025-01-18] Publisher: INFORMS. [Online]. Available: https://pubsonline.informs.org/doi/10.1287/deca.1100.0184

S. A. Ambrose, M. W. Bridges, M. DiPietro, M. C. Lovett, and M. K. Norman, *How Learning Works: Seven Research-Based Principles for Smart Teaching*. John Wiley & Sons, 2010.

H. J. Hartman, "Developing Students' Metacognitive Knowledge and Skills," in *Metacognition in Learning and Instruction: Theory, Research and Practice*, H. J. Hartman, Ed. Dordrecht: Springer Netherlands, 2001, pp. 33–68. [Online]. Available: https://doi.org/10.1007/978-94-017-2243-8_3

D. Dunning, "Chapter Five - The Dunning–Kruger Effect: On Being Ignorant of One's Own Ignorance," in *Advances in Experimental Social Psychology*, J. M. Olson and M. P. Zanna, Eds. Academic Press, Jan. 2011, vol. 44, pp. 247–296. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780123855220000056

R. D. McIntosh, E. A. Fowler, T. Lyu, and S. Della Sala, "Wise up: Clarifying the role of metacognition in the Dunning-Kruger effect," *Journal of Experimental Psychology: General*, vol. 148, no. 11, pp. 1882–1897, 2019, place: US Publisher: American Psychological Association.

J. Leman, C. Kurinec, and W. Rowatt, "Overconfident and unaware: Intellectual humility and the calibration of metacognition," *The Journal of Positive Psychology*, vol. 18, no. 1, pp. 178–196, Jan. 2023, publisher: Routledge _eprint: https://doi.org/10.1080/17439760.2021.1975155. [Online]. Available: https://doi.org/10.1080/17439760.2021.1975155