



Epistemic Consequentialism, Veritism, and Scoring Rules

Marc-Kevin Daoust¹ · Charles Côté-Bouchard²

Received: 26 November 2020 / Accepted: 31 May 2021 / Published online: 10 June 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract

We argue that there is a tension between two monistic claims that are the core of recent work in epistemic consequentialism. The first is a form of monism about epistemic value, commonly known as veritism: accuracy is the sole final objective to be promoted in the epistemic domain. The other is a form of monism about a class of epistemic scoring rules: that is, strictly proper scoring rules are the only legitimate measures of inaccuracy. These two monisms, we argue, are in tension with each other. If only accuracy has final epistemic value, then there are legitimate alternatives to strictly proper scoring rules. Our argument relies on the way scoring rules are used in contexts where accuracy is rewarded, such as education.

Scores are everywhere. Students try to improve their SAT scores for getting into good schools. When buying a house, you need a good credit score, and you might look for a house with a good walking score. Insurance companies measure your health score and your driving score.

According to many thinkers, scores also have their place in the evaluation of *doxastic states*, such as full *beliefs* and *credences* (*degrees of beliefs*). This is a central tenet in the *veritistic epistemic consequentialist* (VEC) research programme. According to VEC, just like we can evaluate a house by giving it a walking score, we can, and should, evaluate doxastic attitudes by giving them *epistemic* scores, i.e., a numbered measure of how well the attitude performs relative to the sole final epistemic goal, namely *truth* or *accuracy*.

In this article, we raise a worry about the viability of a popular brand of VEC (found, e.g., in Greaves & Wallace, 2006; Joyce, 1998, 2009; Pettigrew, 2016). In general, when it comes to determining what scoring system(s) to use to evaluate a

✉ Marc-Kevin Daoust
mk.daoust@live.ca

Charles Côté-Bouchard
charles.cote79@gmail.com

¹ Harvard University, Emerson Hall, 25 Quincy St, Cambridge, MA 02138, USA

² Department of Philosophy, Carleton University, 3A35 Paterson Hall, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada

performance, there are two platitudes that need to be respected. First, which scoring system(s) we should choose depends on which objective(s) we want to reward in evaluating the relevant performances. Second, we should choose the scoring system(s) that best reward the achievement of the objective(s) we want to reward. So, we need to determine both which objective(s) we want to reward and which scoring system does the best job at rewarding the objective(s).

We argue that these two platitudes reveal a tension within this brand of VEC. With respect to the first platitude, VEC identifies *accuracy* as the sole final objective to reward in the evaluation of credences. With respect to the second platitude, VEC typically identifies *strictly proper scoring rules* (SPSR), like the Brier Score, as the sole legitimate way to reward accuracy. At the core of this brand of VEC, then, are two types of *monisms*. One is monism about *epistemic value*: there is not a plurality of final epistemic values. Accuracy is the only final objective or goal in the epistemic domain. All other putative epistemic values can only be instrumental to the final goal of accuracy. The other is monism about a class of *epistemic scoring rules*: there is not a plurality of legitimate systems to measure how well a doxastic state performs relative to the goal of accuracy. Only SPSR are legitimate epistemic scoring rules.

We argue that these two monistic theses are in tension with each other. We begin by highlighting that, in contexts where accuracy matters and is rewarded (like education), a plurality of scoring rules is used. In education, whether a scoring rule is appropriate depends on the *specific objectives* education specialists wish to reward. Of course, education specialists want to reward accuracy. But depending on contextual factors, they also want to reward other things, such as risk-aversion, risk-seeking, or reflection. These additional objectives help them determine which scoring rule is appropriate.

Education, like veritistic consequentialist epistemology, gives a central place to accuracy. However, the lesson from education is that accuracy is not enough to single out a particular class of scoring rules. This lesson reveals a kind of dilemma in which VEC cannot have both monisms at the same time. On the one hand, if SPSR is the only legitimate measure of inaccuracy, then there are, plausibly, other final epistemic values besides accuracy. On the other hand, if accuracy really is the sole final epistemic value, then on some occasions, the recourse to non-strict scoring rules is appropriate.

In Sect. 1, we clarify the two platitudes concerning scoring systems in contexts where accuracy matters and is rewarded, like education. Drawing on these lessons, we present our main argument in Sect. 2: we argue that VEC faces a challenge when it comes to identifying scoring rules that reward the goal of accuracy. VEC is committed to SPSR being the only legitimate system of epistemic scoring. Yet, some non-SPSR are compatible with rewarding accuracy. So, VEC does not exclude scoring rules that are not strictly proper. In Sect. 3, we consider some common *epistemic* arguments for SPSR. We argue that those answers are unavailable to VEC, because they undermine veritism. In Sect. 4, we consider a *pragmatic* justification for excluding non-SPSR as legitimate epistemic scoring systems. This strategy also fails, we argue, because it does not generalize to useful inaccurate credences and accurate credences that are practically detrimental or pointless.

1 Two Platitudes Concerning Scoring Systems

Our argument relies on two platitudes concerning scoring systems. We will begin by saying a few words on these platitudes in contexts where accuracy matters and is rewarded, like education.

Tom finds Sophie's answer sheet to a geography test. Sophie underlined her answers. They go as follows:

1. What is the capital of Belgium? a. Leuven b. Brussels c. Brugge
2. What is the capital of Canada? a. Toronto b. Montreal c. Ottawa
3. What is the capital of the United States? a. Washington b. Miami c. Scranton

Tom knows that Sophie got the right answer to the first question, but got the wrong answer to the second one. She did not answer the third question. He wonders: What is Sophie's score? In order to answer this question, Tom needs to know which *scoring rule* was used for this test.

Scoring rules are functions. They map take an agent's predictions, credences, beliefs, and other doxastic states onto a real number,¹ which represents the agent's score. There is an infinite number of scoring rules. Some are relevant and useful. Others are completely irrelevant. For instance, the following scoring rule is familiar:

Familiar Scoring Rule. For a test of n questions,

- (1) If A gets the right answer to question i , give $1/n$ point to A;
- (2) If A does not get the right answer to question i , give $0/n$ point to A.

But we could create a completely irrelevant scoring rule, such as:

Irrelevant Scoring Rule. For a test of n questions,

- (1) If A gets the right answer to question i , give X points to A where X is the number of moons that a randomly selected planet has;
- (2) If A gets the wrong answer to question i , subtract Y points to A, where Y is the number of blue whales currently in the gulf of Saint Lawrence.

Mathematically speaking, Sophie's score could be anything. However, Tom knows that teachers rarely measure their student's scores with irrelevant scoring rules. So, there must be a subset of scoring rules that are *relevant*, namely, one that rewards the achievement of a valuable objective.

For education specialists, there are different relevant ways to assess a student's score, depending on which valuable objectives the teacher wishes to reward (Budescu & Bar-Hillel, 1993; Lesage et al., 2013). Education specialists care about accuracy, of course. *Ceteris paribus*, an answer sheet with all the good answers should get the best possible score, and an answer sheet with wrong answers only should get a bad score. But the goal of accuracy leaves room for different scoring rules.

¹ With the exception of logarithmic rules—they allow for infinite scores.

For instance, consider the Familiar Scoring Rule. This way of measuring a student's score makes no distinction between wrong answers and the absence of right answers. They both get no point. So, making a guess *dominates* not answering the question, in the sense that students are equal or worse off in every possible world when they do not make a guess (Budescu & Bar-Hillel, 1993, 278).

This means that the Familiar Scoring Rules rewards making guesses, even when one has no clue what the right answer is. In some contexts, guessing and taking risks are welcome. But in other contexts, risk-aversion is preferable. This is why some scoring rules penalize wrong answers (Budescu & Bar-Hillel, 1993; Lesage et al., 2013). Consider the following scoring rule:

Risk-Averse Scoring Rule. For a test of n questions,

- (1) If A gets the right answer to question i , give $1/n$ point to A;
- (2) If A does not get the right answer to question i , give $0/n$ point to A.
- (3) If A gets the wrong answer to question i , subtract $1/n$ point to A.

As Lesage et al., (2013, 189) stress, this scoring rule “penalizes the student for incorrect responses,” and is necessary for discouraging guesses. So, while the Familiar Scoring Rule and the Risk-Averse Scoring Rule both aim at rewarding accuracy, they also comprise secondary objectives, such as encouraging (or avoiding) guessing.

Some “confidence weighting” scoring rules reward self-reflection (Lesage et al., 2013, 190). Depending on how confident they are of their answers, students get differentiated rewards and penalties for each question. For instance, suppose Sophie indicates that she is extremely confident of her answer to question 1. Then, she could get $2/n$ point for a right answer, and $-5/n$ points for a wrong one. These scoring rules encourage students to “know what they know.” When students can distinguish their knowledge from their ignorance, they can maximize their score by giving more weight to their knowledge.

Context could explain why some scoring rules are preferable to others. For instance, suppose you train doctors. We don't want healthcare professionals to think that a wrong answer is equivalent to no right answer. Accordingly, perhaps the Risk-Averse Scoring Rule is preferable for training future doctors. Given that risk-taking is undesirable in healthcare, this scoring rule could be more appropriate for them.

To make a long story short, there is no single way to measure scores. For education specialists, here is how a scoring rule is identified:

1. First, we identify the *valuable objectives* (risk-seeking, risk-taking, self-reflection, etc.) we want to reward.
2. Then, we identify the scoring system(s) that best rewards these valuable objectives.

Since Tom doesn't know what Sophie's teacher wants to reward, he cannot calculate her score. Plausibly, her teacher wants to reward accuracy, but that is compatible with a vast array of scoring rules.

2 Scoring Rules and Veritistic Epistemic Consequentialism

Drawing on the lessons of Sect. 1, we now analyze the extent to which a veritistic epistemic consequentialist can exclude scoring rules that are not strictly proper.

Suppose that Tom knows Sophie's degrees of confidence in the following propositions:

1. She is 80% confident that "Brussels is the capital of Belgium";
2. She is 50% confident that "there is an even number of giant sequoias in Muir Woods".

Tom knows the answers to these questions. This time, he wants to calculate Sophie's *epistemic* score. By epistemic score, we mean the kind of score veritistic epistemic consequentialists are interested in. Given the platitudes outlined in the previous sections, VEC thus needs to answer two questions:

1. First, which *epistemic* objective(s) do they want to reward?
2. Second, which scoring system(s) best rewards these objectives?

2.1 VEC and the platitudes

In general, epistemic consequentialism is the view that the epistemic 'good' is prior to the epistemic 'right'. That is, *conduciveness to epistemic value* determines what we epistemically should believe, and what is epistemically justified and rational to believe; epistemic norms are ultimately about promoting the final epistemic value or values.²

As already indicated, the most common form of epistemic consequentialism is VEC, which takes *truth* or *accuracy* to be the sole final epistemic value to be promoted. Given its commitment to veritism, VEC has a clear answer to the first question: *accuracy* is the sole final epistemic objective to be rewarded by an epistemic scoring system. The second question for VEC is therefore this: which epistemic scoring system(s) best reward(s) accuracy?

Many veritistic consequentialists are committed to a particular answer. To see this, note that for most veritistic consequentialists, an agent's level of accuracy should be measured with *strictly proper scoring rules* (SPSR). Scoring rules are proper if an agent "maximizes the expected score for an observation drawn from the distribution F if he or she issues the probabilistic forecast F , rather than $G \neq F$ " (Gneiting & Raftery, 2007, 359). A scoring rule is strictly proper if the reported probabilities are unique.³

² For more on epistemic consequentialism in general, see Dunn (2015) and Ahlstrom-Vij and Dunn (2018). For criticisms, see Littlejohn (2012, 2015), Berker (2013), and Greaves (2013).

³ Joyce (1998, 2009), Easwaran and Fitelson (2015), Pettigrew (2016), Konek and Levinstein (2017).

Here is an example of how those measures work. Suppose the following propositions are true:

1. Brussels is the capital of Belgium;
2. There is an even number of giant sequoias in Muir Woods.

Suppose Sophie has credences in such propositions. Using SPSR, we can measure how inaccurate her credences are.⁴ Take the Brier score.⁵ When Sophie assigns a credence X in P , her Brier score is measured by the following:

Brier Score:

Suppose A has a credence of X in P . If P is true, the agent's score is $(1-X)^2$. If P is false, the agent's score is X^2 .

The lower Sarah's Brier score is, the better she is from the point of view of accuracy. So, suppose that Sarah has a credence of 0.8 in (1) and a credence in 0.5 in (2). Then, if these propositions are true, her score would be $(1-0.8)^2 + (1-0.5)^2$.

Given the goal of accuracy, SPSR are legitimate reward structures. First, they entail that a perfectly omniscient agent (e.g., with perfectly accurate credences) should get the best possible score. Second, they entail that an agent with credences of 0 in true propositions, and credences of 1 in false propositions, should get the worst possible score. Third, they entail that, all things being equal, an agent who raises his or her credence in a true proposition gets a better score (and, conversely, an agent who lowers his or her credence in a false proposition gets a worse score).⁶

For many veritistic consequentialists, SPSR are the *only* legitimate rules for measuring accuracy. This idea is crucial to the veritistic consequentialist research programme. In particular, SPSR are essential to various vindications of Bayesian norms of rationality, such as Probabilism (Joyce, 1998, 2009; Pettigrew, 2016, chap. 7), Conditionalization (Greaves & Wallace, 2007, Pettigrew, 2016, chaps. 14–15), the Principal Principle (Pettigrew, 2016, chap. 8), and so forth. Without assuming that legitimate scoring rules are strictly proper, most of these results are compromised.

Given all this, many veritistic consequentialists are committed to answering our second question as follows: the epistemic scoring system that best rewards the final epistemic objective (accuracy) is SPSR. If accuracy is the sole final objective to be rewarded and SPSR is the only legitimate epistemic scoring system, then SPSR must be *the* way to reward accuracy. The problem is that this last claim is far from clear.

⁴ See Joyce (1998, 2009), Greaves and Wallace (2006), and Pettigrew (2016) on scoring rules and accuracy-centered epistemology. While accuracy-centered epistemology is concerned with credences, some theories concerned with "full" doxastic states (such as believing, disbelieving and withholding judgment) also assume that having true beliefs is valuable. See, for instance, Easwaran (2015, 826–828) on the epistemic value of true beliefs, false beliefs and judgment suspension.

⁵ See Brier (1950). See Bickel (2007) and Gneiting and Raftery (2007) for comparisons among strictly proper scoring rules.

⁶ See Joyce (2009, 269). These properties roughly reflect his notion of Truth-Directedness. See also Pettigrew (2016, 47–60) on related notions.

2.2 VEC's Trouble with the Second Platitude

Consider the following scoring rules:

Liberal Scoring Rule:

Suppose A has a credence of X in P . If P is true, the agent's score is $(1 - X)^{3/2}$. If P is false, the agent's score is $X^{3/2}$.

Conservative Scoring Rule:

Suppose A has a credence of X in P . If P is true, the agent's score is $(1 - X)^3$. If P is false, the agent's score is X^3 .

As their names suggest, the Liberal Scoring Rule rewards risk-seeking agents, while the Conservative Scoring Rule rewards risk-averse agents. For example, suppose Sophie knows that the objective probability that P is 0.7. The Liberal Scoring Rule entails that assigning a credence of more than 0.7 in P maximizes expected accuracy (in light of her knowledge of the objective probabilities). By contrast, the Conservative Scoring Rule entails that assigning a credence of less than 0.7 in P maximizes expected accuracy (again, in light of her knowledge of the objective probabilities). The Brier Score recommends assigning a credence of exactly 0.7 in P in terms of expected accuracy.

The above scoring rules have the same properties we just discussed. They concur that perfectly omniscient agents get the best possible score, that maximally inaccurate agents get the worst possible score, and that, all things being equal, agents who raise their credence in a true proposition get a better score. This means that, along with SPSR, they reward closeness to perfectly accurate credence functions. So, why not think that, in some contexts, risk-seeking or risk-aversion should be rewarded? Perhaps preachers and stock brokers are epistemically permitted to be risk-seeking, while doctors and engineers are epistemically permitted to be risk-averse. Why not allow for reward structures that match these possibilities?

Think of the discussion, in Sect. 1, on scoring rules used in education. For education specialists, scoring rules reward accuracy. However, we can reward accuracy in different ways, in accordance with different risk profiles. In some contexts, guessing should be discouraged, and in others, it should be encouraged. Someone valuing accuracy can say the same thing. Depending on the context, different risk profiles can be rewarded. The mere goal of accuracy does not rule out this possibility. It is compatible with a plurality of legitimate scoring systems, including the ones that are not strictly proper. If we take the lessons from education seriously, we should not rule out this possibility.

In summary, VEC runs into trouble when answering the following question: Why think that only SPSR are legitimate inaccuracy measures? VEC is committed to SPSR being the only legitimate system of epistemic scoring, yet other systems like the liberal and conservative scoring rules also reward accuracy. Our contention is that, if we take seriously the idea that scoring rules are functions that reward valuable objectives, then truth monism about epistemic value doesn't

necessarily entail that all epistemic scoring rules are strictly proper. It is too weak to rule out other scoring rules.⁷

2.3 Is the Education Analogy Misleading?

Before moving on, we consider a possible worry with what we have said so far. Doing so will help clarify our claim and rule out a possible confusion. One might object that our analogy between epistemic scoring rules and scoring systems in education is misleading on the following grounds: When they think about scoring methods, education specialists are not concerned with determining what it is to be accurate on a test. Rather, they are concerned with finding a scoring method that will best *reveal a student's knowledge*. And these are fairly different things.

More precisely, the objection could be stated as follows: It may be true that there are many ways to measure how we can reveal one's knowledge, but this is not what veritistic consequentialists seek to measure. For them, the issue is not to find a scoring method that best reveals one's knowledge or, more relevantly, one's credences. Instead, the issue is to measure *how accurate* one's credences are. This is different from the issue of determining what's in people's head, as it were. So, the analogy between educational scoring and epistemic scoring is misleading. The fact that there are many ways to measure one's knowledge does not entail that there are many ways to be accurate.⁸

To respond, we need to start with a clarification. The question we are concerned with is *not* "what is it to be accurate?". We know, for instance, that having a credence of 0.8 in a true proposition is more accurate than having a credence of 0.7 in the same true proposition. This is a fact about the nature of accuracy. Rather, the question we are concerned with is this: How should we *score* credences with respect to accuracy? And our point is that accuracy alone *underdetermines* the choice between scoring systems. Even if we agree that accuracy is the fundamental aim, and even if we agree on what it is to be accurate, this does not fully settle the question of which scoring system we should choose.

And this is where the analogy with scoring rules in education is helpful. In an education setting, we can hold fixed the knowledge that the student really has. We can assume, in other words, that there's an objective fact about that. But that fact alone is not enough to settle the choice between different scoring systems. In order to select one scoring system over another, we must make various choices, and accuracy underdetermines these choices. Additional aims guide these choices. The same goes with accuracy and scoring systems for credences. Accuracy alone underdetermines the choice between various scoring systems. To pick one system rather than another, we must make various choices, and the fundamental criterion of the credence's accuracy is neutral about these choices.

⁷ And of course, perhaps only SPSR are legitimate inaccuracy measures. But then, the lesson would be that we should reject veritistic value monism.

⁸ We thank an anonymous referee for raising this worry.

3 Veritism and Arguments for Strict Propriety

In the previous section, we have argued that veritism is compatible with scoring rules that are not strictly proper. In this section, we confirm this by analyzing three prominent epistemic arguments for strict propriety. We argue that they all violate veritistic monism (Note: Some of our objections are not new. But in order to provide a comprehensive overview of the issue, we present them anyway).

3.1 Order-Sensitivity

An initial motivation in favour of strictly proper scoring rules comes from D'Agostino and Sinigaglia (2010). They think that scoring rules should satisfy what they call Order-Sensitivity. In order to understand Order-Sensitivity, consider the following case:

	x	y	y'
E1	0.3	0.6	0.9
E2	0.6	0.9	0.6

E1 and E2 are propositions, and x, y and y' are judgments concerning E1 and E2. For instance, x thinks that there is a 30% chance that E1, and that there is a 60% chance that E2. The difference between y and y' is that their credences in E1 and E2 are “swapped”: the order of their credences in E1 and E2 is reversed.

D'Agostino and Sinigaglia think that the distance between x and y is greater than the distance between x and y'. The reason for this, they argue, is that x and y agree *qualitatively* on their judgments concerning E1 and E2. That is, x and y concur that E2 is more likely than E1. They say:

It seems natural to judge that the distance between x and y' be greater than the distance between x and y: for x and y agree *qualitatively* on the degree of belief they assign to the events E1 and E2, since they both express a higher degree of belief in E2 than in E1, and so any inversion of this qualitative judgement – as in y' – must be regarded as increasing the distance between the belief-vectors; intuitively, y' disagrees with x more than y. Moreover, it seems reasonable to assume that the additional distance between y' and x grows monotonically with the difference between the inverted values (0.3 in this example). This also appears to be intuitively sound (D'Agostino & Sinigaglia, 2010, 101).

Not all scoring rules would say that the distance between x and y is less than the distance between x and y'. For instance, the absolute distance between x and y is the same as the absolute distance between x and y'. When a scoring rule says that y is closer to x than y', it is sensitive to the qualitative agreement between x

and y . In short: These scoring rules satisfy what D'Agostino and Sinigaglia call Order-Sensitivity.⁹

D'Agostino and Sinigaglia think that Order-Sensitivity is a legitimate constraint on scoring rules. A good scoring rule measuring the accuracy of our credences should satisfy it. And when combined with other common assumptions,¹⁰ it excludes scoring rules that are not strictly proper (see, e.g., D'Agostino & Dardadoni, 2009). Thus, Order-Sensitivity is a crucial step in D'Agostino and Sinigaglia's vindication of SPSR.

Does this argument vindicate strict propriety? And more importantly, is their argument compatible with thinking that strictly proper scoring rules reward accuracy *along with something else*? First, note that D'Agostino and Sinigaglia's argument relies primarily on intuitions. For them, it is *intuitive* to regard the distance between x and y' as greater than the distance between x and y . However, there are other intuitions that conflict with SPSR. For instance, consider the following case:

	x	z	z'
E1	1	0.6	1
E2	1	0.5	1
E3	1	0.4	0.10

Suppose x is the infallible prediction. For instance, x assigns the most accurate credence in E1, E2, and E3. It is intuitive to regard z' as more accurate than z (at least, we find it intuitive to regard z' as more accurate than z , and people around us too). If z and z' were two competing weather channels, we would watch z' . However, this intuitive reaction conflicts with some plausible strictly proper scoring rules. For instance, relative to the Brier score, z gets a better score than z' .¹¹

Intuitions concerning the accuracy of judgments and predictions can point in many different directions. Some of our intuitions may favour SPSR, but other intuitions may favour alternatives to SPSR. Here, the lesson is either one of two things: (1) In these debates, we have various intuitions that point in different directions, and thus are not a reliable guide, or (2) provided that our intuitions are a reliable guide, we have a reason to think that there are legitimate alternatives to SPSR (since some of our intuitions conflict with SPSR).

Second, even if our intuitions were reliable and favoured Order-Sensitivity, we would still lack a reason to think that *accuracy alone* explains this. Suppose, in accordance with D'Agostino and Sinigaglia, that we regard y as closer to x than y' . How can we explain this? Perhaps our intuitions are best explained by dispositions or preferences that go beyond the mere goal of accuracy. That is, not only accuracy should be rewarded, but also some specific ways to be accurate.

⁹ See D'Agostino and Sinigaglia (2010, 102) for a formal definition of Order-Sensitivity.

¹⁰ The assumptions and properties are: One-Dimensional Value-Sensitivity, Sub-vector Consistency, Permutation Invariance and Replication Invariance. See D'Agostino and Sinigaglia (2010, 100–104).

¹¹ Relative to the Brier score, the predictions made by z get a total score of 0.77, while the predictions made by z' get a total score of 0.81. And recall that agents are trying to *minimize* their Brier score.

Recall that the absolute distance between x and y is “less extreme” (i.e., less close to values of 0 and 1) than the distance between x and y' :

	x	y	y'	Absolute distance between y and x	Absolute distance between y' and x
E1	0.3	0.6	0.9	0.3	0
E2	0.6	0.9	0.6	0.3	0.6

Given this, the reason why our intuitions favour Order-Sensitivity could be that, all things being equal, we wish to avoid extreme (absolute) distances. That is, provided that x is a reference point for accuracy (i.e., that being closer to x is being more accurate), our intuitions suggest that it is better to have two moderately inaccurate attitudes in E1 and E2 than one perfectly accurate attitude in E2 and one highly inaccurate attitude in E2. In other words, in addition to valuing accuracy, we also value certain ways to be accurate. Those who think that y is closer to x than y' give more importance to the goal of avoiding inaccurate credences than to the goal of having accurate credences. So, D’Agostino and Sinigaglia’s argument is compatible with thinking that veritistic consequentialism does not vindicate SPSR *on its own*. The putative reasons why we favour Order-Sensitivity could go beyond mere accuracy.¹²

3.2 Immodesty

According to Gibbard (2007) and Joyce (2009), the credences warranted by scoring rules must be compatible with some facts about epistemic rationality. Specifically, the credences warranted by scoring rules should not conflict with the *immodesty* requirement. Joyce writes:

Modest credences, it can be argued, are epistemically defective because they undermine their own adoption and use.... If, relative to a person’s own credences, some alternative system of beliefs has a lower expected epistemic disutility, then, by her own estimation, that system is preferable from the epistemic perspective. This puts her in an untenable doxastic situation.... This is a probabilistic version of Moore’s Paradox. (Joyce, 2009, 277)

Joyce here claims that rational agents are immodest, in the sense that they take their own credences and standards to be among the most accurate (or truth-conducive) ones available to them. The requirement of immodesty is a requirement of coherence in the broad sense. It governs the relationship between an agent’s doxastic states and her beliefs concerning such states (see also Lewis, 1971; Elga, 2010).

It is incoherent for agents to be modest. If one thinks that some doxastic states or standards are epistemically superior, one should feel some pressure to change one’s doxastic states and standards. Entertaining doxastic states and standards that

¹² We thank an anonymous referee for inviting us to discuss this argument.

one takes to be less than optimal entails some sort of incoherence or self-distrust. In view of the foregoing, many philosophers think that agents are epistemically required to be immodest.¹³ Relatedly, many recent philosophical arguments presuppose that agents are epistemically required to be immodest.¹⁴

Immodesty is used to rule out scoring rules that are not strictly proper.¹⁵ Joyce writes:

Some credences do seem uncontroversially rational, and so... [we can] rule out various candidate scoring rules. Consider the exponential scores with $z \neq 2$ [like the liberal or the conservative scoring rules].... Normally, it seems fine to align one's credences for the tosses of a fair die with the uniform probability over the six sides. However, with the exception of the Brier score, all exponential scores make such credences modest. (Joyce, 2009, 278)

Very roughly, Joyce uses immodesty, a coherence requirement (or, more specifically, a requirement prohibiting self-distrust), as a constraint on available scoring rules. The liberal scoring or the conservative scoring rule could make it modest, for an agent, to entertain credences that are clearly permissible or required. Joyce thinks this is problematic. So, he excludes them *on this ground*—that is, these scoring rules are ruled out *because* they make it incoherent, for agents, to have some permissible credences. Hence, immodesty is used as a constraint on legitimate scoring rules.

However, there are known worries concerning the claim that agents fall under an obligation to be immodest. For instance, according to Christensen (2013), epistemic modesty can be permitted in cases where various epistemic ideals (e.g., ideals of coherence, good reasoning and evidence-responsiveness) conflict with each other. Relatedly, DiPaolo (2018) and Daoust (2020) argue that modesty can be part of the best non-ideal option available to agents.

Also, why might veritists want to reward immodesty and, more generally, coherence? Joyce may be right that being immodest is epistemically valuable. Perhaps self-trust, or coherence more generally, is epistemically valuable. But if VEC's value monism is right, then it can only be epistemically valuable for the sake of the sole final epistemic value of accuracy. Given veritism, self-trust and coherence can be epistemically good only insofar as it promotes accuracy. So, veritists should want

¹³ Lewis (1971) and Elga (2010) suggest that modest methods or standards cannot be completely trusted, and thus display a kind of internal failure. See also Hedden (2015), Horowitz (2014) and Schoenfield (2014). Mayo-Wilson and Wheeler (2016) argue that, in order to accommodate scoring rules for imprecise credences, immodesty ought to be replaced with a weaker principle.

¹⁴ For more on this issue, see e.g., Hedden (2015), Horowitz (2014), Schoenfield (2014), and Li (2016, 9–10).

¹⁵ Along with other assumptions and principles. For instance, Joyce's vindication assumes that the Principal Principle is correct. This principle says that an epistemically rational agent's credence ought to reflect his or her knowledge of the objective probabilities. Combining Immodesty, the Principal Principle and other common assumptions leads to strict propriety. However, Hájek (2009, 247–249) raises some worries concerning the recourse to the Principal Principle in Joyce's argument. Among other things, Hájek is worried that there can be conflicts between the Principal Principle and what Joyce calls Minimal Coherence.

their scoring system to reward immodesty, self-trust or coherence just in case doing so were instrumental to rewarding accuracy.

But this does not seem to be Joyce's story. Immodesty and coherence are presented as worth rewarding not because they promote accuracy, but because they are epistemically good in their own right; because they are clear requirements of epistemic rationality and we do not want to reward epistemic irrationality. So, the justification is not that rewarding immodesty is necessary to reward accuracy, but rather that it is necessary to avoid rewarding incoherent credences. This suggests a value-pluralistic story according to which the goals we need to reward include not only accuracy, but also coherence or self-trust.¹⁶

Here are two replies to this worry. First, perhaps we can use requirements like immodesty to figure out the nature of inaccuracy measures. For instance, immodesty could be used in an inference to the best explanation pertaining to inaccuracy measures, as in the following¹⁷:

(P1) Measuring inaccuracy with SPSR allow rational agents to take the rational credence in P and be immodest;

(P2) Measuring inaccuracy with alternatives to SPSR can put rational agents in a difficult position: If they take the rational credence in P, they could have to be modest about their credence in P.

(P3) The best explanation of (P1) and (P2) is that inaccuracy is measured with SPSR.

(C) So, inaccuracy is measured with SPSR.

However, even if we assume that this type of reasoning carries weight, there is an explanatory gap between (P1)–(P2) and (P3). That alternatives to SPSR create problems for rational agents should lead us to infer something about the relationship between rationality and scoring rules. A weaker inference from (P1)–(P2) is this:

(P3)* The best explanation of (P1) and (P2) is that, if agents value rationality, then they measure inaccuracy with SPSR.

(C)* So, if agents value rationality, they measure inaccuracy with SPSR.

(C)* is compatible with our argument. Traditionally, there are different dimensions associated with rationality: reason correctly,¹⁸ avoid incoherent combinations

¹⁶ There could, of course, be veritistic justifications of immodesty. But these putative vindications are unheard of. Here, we limit ourselves to the known justification for immodesty, based on coherence and self-trust. Should new justifications be offered for this requirement, we will be able to analyze them.

¹⁷ There are similar strategies in print. Horowitz (2014), for instance, is concerned with the relationship between Immodesty and permissivism. That is, she tries to figure out whether uniqueness best explains Immodesty. We thank a referee for inviting us to discuss this point.

¹⁸ See, e.g., Broome (2013, chaps. 12–16).

of attitudes,¹⁹ form and revise attitudes in accordance with your evidence,²⁰ optimize expected utility or accuracy,²¹ etc. In (C)*, SPSR could be a consequence of dimensions of rationality that are tangential to accuracy. For instance, the reason why rational agents measure inaccuracy with SPSR could be that rationality rewards accuracy *and* coherence. Once again, we lack a reason to think that accuracy *on its own* justifies SPSR.

Second, one could reply that this story is still compatible with veritism since rewarding coherence (or self-trust) is, in fact, necessary to reward accuracy.²² For instance, perhaps requirements of coherence are explained away by other requirements, such as maximizing expected accuracy. If this is right, then appeals to immodesty would remain true to veritism.

But this alleged relationship between accuracy and coherence is contentious. There is growing concern that coherence is not a necessary condition of various epistemic norms and goals. Some philosophers have argued that coherence can conflict with evidence-responsiveness (Coates, 2012; Lasonen-Aarnio, 2018; Worsnip, 2018) and some substantive rational requirements (see, e.g., Kiesewetter, 2017, 254, 258). So, coherence is not entailed by various epistemic norms.

What about norms of (expected) accuracy? In the literature on veritistic norms for full belief, many philosophers have argued that coherence and the maximization of expected accuracy can come apart. Just think of the Lockean consequentialist picture of full belief.²³ For Lockean consequentialists, the maximization of expected accuracy sometimes entails that, while an agent maximizes expected accuracy in believing P_1, P_2, \dots, P_n , he or she maximizes expected accuracy in disbelieving the conjunction (P_1, P_2, \dots, P_n) . In technical terms, maximizing expected accuracy is sometimes incompatible with *closure principles*. Beliefs that are not closed under conjunction cannot be true simultaneously, and thus display a kind of incoherence.²⁴ So, expected accuracy maximizers are not necessarily coherent.

Now, perhaps the Lockean consequentialist picture of belief is false. This example is simply meant to show that some expected accuracy principles can conflict with coherence principles. So, the claim that rewarding coherence is necessary to reward accuracy cannot be taken for granted. It is a live possibility that coherence can conflict with maximizing (expected) accuracy.

¹⁹ See, e.g., Broome (2013, chap. 8–9), Worsnip (2018).

²⁰ See, e.g., Kiesewetter (2017), Lord (2018). Kiesewetter and Lord refer to epistemic reasons, not evidence. But Kiesewetter thinks that one's reasons are "evidence-relative."

²¹ See, e.g., Wedgwood (2017). But Wedgwood refers to correctness instead of accuracy.

²² We take modesty to be a species of incoherence, namely, self-distrust. See, e.g., DiPaolo (2018), Elga (2010) and Schechter (2013).

²³ See Dorst (2017) and Easwaran (2015).

²⁴ This is a common observation in the literature on the lottery paradox. See, e.g., Nelkin (2000) and Hawthorne and Bovens (1999).

3.3 Calibration and Decomposition

Another justification for VEC's epistemic scoring monism—for the claim that SPSR are the only legitimate inaccuracy measures—relies on calibration, and comes from Pettigrew (2016, sec. 4.3). Pettigrew begins by quoting Ramsey, who says:

Granting that [an agent] is going to think always in the same way about all yellow toadstools, we can ask what degree of confidence it would be best for him to have that they are unwholesome. And the answer is that it will in general be best for his degree of belief that a yellow toadstool is unwholesome to be equal to the proportion of yellow toadstools that are unwholesome. (Ramsey, 1931, 195)

Pettigrew then analyzes the meaning of “best” in Ramsey’s quote. In what sense is it “best” to proportionate one’s credences in “a yellow toadstool is unwholesome” to the observed frequencies (e.g., the observed proportion of yellow toadstools that are unwholesome)? Surely, epistemically speaking, an agent would be better off if he or she predicted perfectly which yellow toadstools are wholesome and which are not (Pettigrew, 2016, 58).

However, there is a different sense in which it would be “best” to proportionate one’s credence to the observed frequencies. The difference can be explained in terms of reference class. If each individual toadstool has its own reference class, then the observed frequencies and the perfect credence function will coincide. But if a reference class contains many yellow toadstools, it could be ideal for an agent to proportionate one’s credences to the observed frequencies in this reference class. As Pettigrew says, “on Ramsey’s account, the ideal credence function depends on the reference classes into which we partition the set F of propositions to which the agent assigns credences.” (Ibid., 58) In other words, with respect to a reference class, one’s credences should be calibrated.

Pettigrew thinks that legitimate scoring rules should satisfy calibration.²⁵ He wishes to make sense of the intuition that “credences are better the closer they are to being well calibrated.” (Ibid., 63) When combined with other desiderata on legitimate scoring rules,²⁶ we get the result that only SPSR are legitimate. (Ibid., 65–66).

But just like the arguments from order-sensitivity and immodesty, this appeal to calibration does not remain faithful to veritistic value monism. To see this, consider Pettigrew’s response to the following objection against calibration. Suppose Bonnie and Clyde know that 50% of the toadstools in their garden are wholesome. There are two mushrooms in their garden: C_1 and C_2 . They both think that only C_1 and C_2 are part of the relevant partition class. As it happens, C_1 is wholesome, and C_2 is not. Bonnie has a credence of 0.5 in “ C_1 is wholesome” and a credence of 0.5 in “ C_2 is wholesome.” Clyde has a credence of 0.9 in “ C_1 is wholesome” and a credence of 0.2 in “ C_2 is wholesome.” Bonnie has calibrated credences, while Clyde doesn’t.

²⁵ Or, more specifically, Decomposition (Ibid., 63).

²⁶ These other desiderata are Alethic Vindication, Perfectionism, and Continuity, and Additivity. See Pettigrew (2016, 47–60).

However, Clyde's credences are more accurate. So, we are faced with a problem: Calibration seems, on some occasions, to conflict with truth-directedness. If we merely want to reward accuracy, we should not necessarily reward calibration.

Pettigrew replies that calibration offers a specific interpretation of how we should pursue accuracy. Agents primarily want to be accurate, but they may only pursue this goal by satisfying calibration. He thinks that both desiderata (calibration and accuracy) do not necessarily conflict with each other, because we can evaluate an agent's calibrated credence function by its proximity to the perfectly accurate credence function. He says:

We might preserve the intuition behind calibrationism and the intuition behind truth-directedness by saying that the accuracy of a credence function at a world is partly determined by its proximity to its well calibrated counterpart, but partly by the proximity of its well calibrated counterpart to the omniscient credence function. This retains both intuitions: the truth-directedness intuition is that proximity to the omniscient credence function is important; the calibrationist intuition is that proximity to the well calibrated counterpart is important. (Ibid., 63)

To say this, however, is to make an epistemic value claim that does not trace back to the final value of accuracy. As Pettigrew notes himself, calibration accuracy is logically stronger than accuracy *tout court* (Ibid., 63). If all we care about is accuracy, we have no particular reason to prefer the stronger interpretation of accuracy in terms of calibration.

To be clear: epistemically speaking, rewarding calibration can make perfect sense. And perhaps we have strong intuitions in favour of calibration. Our contention is that invoking the need to reward this desideratum goes against veritistic monism. Calibration is a secondary objective that restricts how we should pursue accuracy. Some veritists might not want to reward calibration.

Compare with an example discussed in Sect. 1: We want doctors to be accurate. However, we don't want them to guess. This is why, when training doctors, we go for scoring rules that reward risk-averse accuracy. These scoring rules reward two objectives: accuracy *and* risk-aversion (or whatever final goal is promoted by doctors' being risk-averse). Similarly, Pettigrew is looking for a scoring rule that rewards accuracy, but also penalizes agents when they violate calibration. This is why Pettigrew goes for calibration accuracy. But again, this simply means that scoring rules should reward two objectives: accuracy and calibration (or whatever final goal is promoted by calibrating). Talk of "risk-averse accuracy" is similar to talk of "calibration accuracy": In both cases, this simply means that two independent objectives are rewarded. In both cases, we wish to reward accuracy *along with something else*.²⁷

²⁷ Note that this worry is not entirely new. Levinstein (2017b, sec. 2.3) has argued that Pettigrew's commitment to calibration can be interpreted in two ways: either (1) he endorses some sort of epistemic value pluralism, or (2) he does not remain faithful to the desideratum of truth-directedness (by introducing a constraint that is outside the scope of truth-directedness). In response to Levinstein's argument, Pettigrew reiterates that calibration acts as a secondary goal, and provides a specific interpretation of how to pursue the goal of accuracy (Pettigrew, 2017, 11). This responds to the second interpretation proposed by Levinstein, namely, (2). But Pettigrew's response is compatible with (1), namely, epistemic value pluralism.

3.4 Taking Stock

We have examined three prominent arguments, from proponents of VEC, for SPSR being the only legitimate epistemic scoring system. We have argued that those justifications of SPSR conflict with veritistic monism.

As we have seen at the outset, which scoring system(s) we should choose depends on which value(s) we aim to reward. VEC is committed to accuracy being the only epistemic goal to be rewarded for its own sake. Any other epistemic goal should be rewarded just in case it is necessary to reward accuracy. The problem is that the arguments for SPSR-based monism invoke the need to reward goals that are not only distinct, but also independent from accuracy.

To be clear: It is an interesting phenomenon that many different desiderata converge towards strict propriety. But none of the above arguments show that, if we merely care about accuracy, there cannot be, in some contexts, legitimate reward structures for risk-seeking or risk-averse agents. These arguments suggest that strict propriety makes sense in many contexts, but they do not rule out the possibility of “local” alternatives to strictly proper scoring rules.

4 Levinstein’s Pragmatist Vindication of SPSR

The vindications of SPSR-based monism we have seen so far have been epistemic in character. They invoked a scoring rule’s capacity to reward goals that are recognizably epistemic. This, however, was part of the problem. Those epistemic vindications did not trace back to the need to reward accuracy, thereby undermining VEC’s other monistic thesis that views accuracy as the sole final epistemic value.

In light of this, could veritists instead invoke *pragmatic* reasons for excluding non-SPSR? This could be a promising strategy since it might allow VEC to justify its SPSR-based monism without invoking any other *epistemic* goal besides accuracy, thereby leaving its epistemic value monism unaffected. After all, the claim that accuracy is the sole final epistemic value is compatible with the existence of non-epistemic final values.

One recent example of this kind of strategy comes from Benjamin Levinstein (2017a). According to him, we can derive SPSR from pragmatic concerns (or from the assumption that the epistemic value of accuracy is grounded in practical value).

In the remainder of this section, we analyze his argument, and identify some of its limits.

Footnote 27 (continued)

If, on top of accuracy, we wish to reward an independent secondary objective, we must assume that it is valuable independently of accuracy.

4.1 Pragmatism and Veritism

Levinstein is a pragmatist, in the sense that he takes all value to be related to practical success. Yet, he thinks that accuracy has final epistemic value. Some might think that these views are inconsistent with each other. Let us clarify how one can be a pragmatist and also a veritist about epistemic value.

There are many *domains* that imply standards or norms. In addition to *epistemic* norms, there are also norms associated with morality, prudence, fashion, etiquette, finance, tradition, games, etc. Not all of those domains are equally important, of course. But they all set standards relative to which certain behaviours count as permitted, forbidden, obligatory, correct, incorrect, good, bad, and so on.

Veritists think that accuracy (or truth) has final *epistemic* value. Minimally, this means that, relative to the epistemic domain, accuracy is valuable. But this is compatible with saying that there is a more general domain, like the practical one, from which all the relevant domain-relative values are derived. This is Levinstein's view. He thinks that all domain-relative values that matter can be derived from a concern for practical success. Whether something has *epistemic* value depends on whether it ultimately promotes accuracy. But whether we ought to promote epistemic value depends, ultimately, on whether doing so would promote practical value. Such a pragmatist might then argue that epistemic value is worth promoting because doing so is, in general, necessary for practical success. As Levinstein writes, "we are in agreement with the pure epistemic utility theorist that the only epistemic good is accuracy. ... We simply disagree about the ultimate source of this value" (Levinstein, 2017a, 615).

So, the reason we address Levinstein's argument is that he thinks, in accordance with veritistic consequentialism, that accuracy has final epistemic value. However, he thinks that we can explain the epistemic value of accuracy in terms of practical goals. And perhaps taking a closer look at these practical goals can tell us why epistemic scoring rules are strictly proper.

4.2 A Brief Summary of Levinstein's Argument

In this section, we briefly present Levinstein's argument for strictly proper scoring rules.

First pass: Levinstein starts by assuming that rational agents will try to minimize their expected losses (2017a, 622, 624). In decision theory, we are familiar with cases where an agent's evaluation of the expectancies will depend on her credence in various events. For instance, if your credence in the proposition "it will rain today" is low, you will probably judge that not bringing your umbrella to work will maximize expected utility. By way of contrast, if your credence in the proposition "it will rain" goes up, this might affect your judgment that not bringing your umbrella to work will maximize expected utility. This is so, because the expected utility of bringing an umbrella to work will be higher. So, having accurate credences often has practical implications.

In technical terms, we could say that if agents have accurate credences, they are more likely to be on the right side of the *cutoff* (ibid., 626, 628). A cutoff is a credence q in P over which an agent will perform an action (that maximizes expected utility). Being on the right side of the cutoff means performing the best action. For instance:

- Suppose it will rain tomorrow. If Alice has a credence of more than 0.6 in “it will rain tomorrow,” she will bring her umbrella. Then, 0.6 is the cutoff, and Alice will be on the right side of the cutoff if she has a credence of more than 0.6 in “it will rain tomorrow” (on this side of the cutoff, she will bring her umbrella). Higher credences in the proposition “it will rain tomorrow” are more likely to put Alice on the right side of the cutoff. Given that it will rain tomorrow, raising her degree of confidence in this proposition increases her accuracy. So, accuracy makes it more likely that she will be on the right side of the cutoff.
- Suppose it will not rain tomorrow. Then, Alice will be on the right side of the cutoff if she has a credence of less than 0.6 in “it will rain tomorrow.” Again, accuracy makes it more likely that she will be on the right side of the cutoff.

Naturally, perfectly accurate agents would always make decisions that are on the right side of the cutoff.

Second pass: Given that agents want to minimize expected disutility (or maximize expected utility), accurate credences are useful when they are relevant to problems that have a cutoff (Levinstein calls them “ q -problems”). Accordingly, we can determine the expected utility of having accurate credences. Since we want to minimize the expected loss, we need to know both “how high the stakes are expected to be given that a q -problem is faced and how likely Alice thinks it is that she will face a q -problem” (ibid., 629).²⁸

This, of course, presupposes that the cutoff (i.e., the credence q) is known. We do not always know this. For generality, Levinstein assumes that an agent could face a q -problem, where $q \in [0,1]$. Then, what will matter in such a case is the *probability density* of facing a q -problem multiplied by the expected stakes associated with that problem.²⁹

Final pass: Then, Levinstein argues that, since rational agents try to minimize their expected losses, they measure the inaccuracy of their credences with strictly proper scoring rules. His strategy is based on Schervish’s representation theorem. Very roughly, he argues that if the function of probability density of facing a q -problem multiplied by the expected stakes associated with that problem³⁰ is strictly positive almost everywhere (e.g., not always zero), then the loss function that returns how much A will gain (or lose) given A ’s choice of credences will be strictly proper. See Levinstein (2017a, 630) and Schervish (1989) for details. In other words, given

²⁸ Formally: Agents want to optimize the function given by $E(W|q) \cdot \Pr(q)$. See (ibid., 629).

²⁹ Formally: Agents want to optimize the function given by $p(q) \cdot E(W|q)$. See (ibid.).

³⁰ Formally: The function given by $p(q) \cdot E(W|q)$. See the previous footnotes.

our practical reasons to value accuracy, it should be measured with strictly proper scoring rules.

4.3 Two Worries with Levinstein's Argument

We grant that Levinstein's argument can apply to many situations. However, this is not enough to show that an agent's level of inaccuracy must *always* be measured with SPSR.

First, Levinstein focuses on cases where accuracy is a means for being on the right side of the cutoff. As discussed in Sect. 4.2, a cutoff is a credence q in a proposition P over which an agent will perform an action (that maximizes expected utility). Agents are on the right side of the cutoff when their decision is optimal.

In most cases, accuracy is instrumental for making the best decision (or being on the right side of the cutoff). However, there are cases in which this assumption is not true. Suppose that Alice is very sick. She would be greatly comforted by the thought that she will survive. Perhaps it is very unlikely that she will survive. So, having an inaccurate credence in the proposition "I will survive" is the best decision she can make. Note that Alice's degree of inaccuracy in the proposition "I will survive" doesn't make it more likely that she will be on the wrong side of the cutoff. She will find comfort in thinking that she will survive, no matter how inaccurate this credence is.

Insofar as accuracy is conducive of good decisions, Levinstein's argument is plausible, and strict propriety makes sense. But the argument does not generalize to the (not uncommon) cases in which accuracy is independent of good decisions, i.e. cases in which accuracy would lead to bad decisions or where inaccuracy would lead to good decisions. Accordingly, Levinstein's pragmatist vindication of SPSR does not apply to cases in which having inaccurate doxastic states serves an agent's goals. This is compatible with there being some contexts in which an agent's inaccuracy could be measured with scoring rules that are not strictly proper.

In addition to detrimental truths and useful falsehoods, the possibility of practically *pointless* or *trivial* truths is a problem for Levinstein's argument. The kind of true belief (or accurate credence) we here have in mind is about a subject-matter or topic that has absolutely no connection to your practical life. It is about a topic that is utterly irrelevant and unrelated to any of your pragmatic goods. Think of topics like what goes on in some nearly empty corner of the universe, the lives of some insect from the distant past, and the like. The best examples of pointless truths are also extremely complex, detailed, and precise. As Marian David points out:

'[A]ll truths' comprises not only each and every trivial truth of the 'how many threads in my carpet'-variety, but also every conjunction of every two truths I already believe (trivial or non-trivial), including, of course, complex conjunctions whose conjuncts are themselves conjunctions of truths I already believe (trivial or non-trivial); it also comprises every disjunction of (trivial or non-trivial) truths I already believe with any propositions you like; and lots of other redundant 'garbage' of this sort. (David, 2005, 296–298)

Some true beliefs are not just about irrelevant matters, they are also about highly complex propositions that logically connect any number of those countless irrelevant beliefs we can form. At first sight, there is no reason to think that having accurate doxastic states in these propositions is practically valuable.³¹

Levinstein's argument doesn't generalize well to pointless truths. Take the following example. What do you get when you take the number of blades of grass in the eight largest parks in Belarus at 10:55 pm on October 1st, 2011 and divide it by the number of hairs on the left arm of the second oldest orangutan at the London zoo in 1993? There is a truth about that, but it is hard to see what practical value there could be in believing it. It is patently clear that there is nothing at stake in having a true belief in this proposition.

One natural objection is that even those obscure complex beliefs are worth promoting, practically, because they *could*, in principle, turn out to be useful down the line. We never know whether an accurate credence in P (or a true belief in P) will be useful to us in the future (Levinstein, 2017a, sec. 3.5).

However, we can imagine cases in which we *know* that an accurate credence in P is pointless. Suppose you know for sure that you will die in the next 10 s or that your recent memory and beliefs will be wiped out. You can't record any information and you can't communicate with anyone. Perhaps you jumped off a plane without a parachute naked in the middle of Antarctica while ingesting a high dose of deadly poison. Before dying, you could acquire an accurate credence concerning the number of giant sequoias in Muir Woods. You really don't care about sequoias. Moreover, you know that an accurate credence in "there is an even number of giant sequoias in Muir Woods" won't satisfy your sense of curiosity, won't stimulate your intellect, won't help you reach new interesting conclusions, won't be useful to the members of your epistemic community, and so forth. Plausibly, you know that accuracy concerning the number of giant sequoias in Muir Woods is pointless for you (given your situation). The upshot is that in some circumstances, you *can* come to know that an accurate credence in P is entirely pointless for you.

Now, suppose agents know that an accurate doxastic state in a given proposition is entirely pointless. Accordingly, there is nothing at stake with having an accurate credence in this proposition.³² In Levinstein's framework, this means that the probability density of facing a q-problem multiplied by the expected stakes associated

³¹ Note that our argument commits us only to the existence of *practically* pointless truths, i.e., true propositions such that believing them would not promote any *practical* value such as the well-being of the believer's or of others in the believer's community. We do not need the closely related, but stronger claim that some truths are such that believing them would not promote *any* genuine value, not even *non-practical* value. For instance, what we say is compatible with the claim that although some true beliefs are practically pointless, they are worth having because every true belief is valuable for its own sake (just in virtue of being true). So, all we are saying is that whether or not true belief itself is a (non-practical) final good, it is not necessarily valuable for the sake of some practical good. For more on this issue, see Lynch (2004), David (2005), Kvanvig (2008), Hazlett (2013), Whiting (2013), Wrenn (2017), and Côté-Bouchard (2017).

³² Given Levinstein's notation, this amounts to knowing that $W = 0$.

with that problem will be equal to 0.³³ Why? Because agents know that nothing is at stake, and so, the expected stakes are null.

If the function of probability density of facing a q-problem multiplied by the expected stakes associated with that problem is equal to 0, the final pass of Levinstein's argument based on Schervish's representation theorem doesn't work. His argument works if the function is strictly positive almost everywhere. In natural language: The function should not be equal to 0 everywhere. This condition is violated in cases where a proposition is entirely pointless. Thus, in cases where agents know that nothing is at stake, or that a proposition is practically pointless, the pragmatist vindication of SPSR is unsuccessful.

So, at best, Levinstein's argument provides a *partial* vindication of SPSR. His argument is compatible with the claim that, in cases where there are no practical stakes (e.g., having an inaccurate credence in P won't matter from a practical point of view), there are alternatives to SPSR.

Levinstein acknowledges that there are pointless propositions. Formally, this possibility is represented in his framework by $W=0$. He could even grant that agents can *know* that a proposition is pointless. However, he also notes that agents are often *uncertain* of the practical stakes associated with having accurate credences in various propositions. So, in most contexts, agents assume that W can take value 0, but that it could also take value 2, 10, 100, and so forth. If agents think that W can take another value than 0, the argument succeeds. Indeed, as Levinstein notes: "if Alice leaves open the possibility (however remote) of facing any q-problem whatever, then she will use a strictly proper scoring rule to evaluate her credence." (Levinstein, 2017a, 630).

So, Levinstein could argue that all he needs for his argument to succeed is that agents are *not fully certain* that the proposition is practically pointless (i.e., they have a credence of less than one in "this proposition is practically pointless"). The claim would then be that rational agents can never have credence 1 that a proposition is pointless.

But going this route would have controversial consequences. For starters, consider the orthodox Bayesian claim that an agent's evidence consists in the propositions to which she assigns credence 1.³⁴ While some philosophers think that some evidence deserves a credence of less than 1, few would argue *none* of our evidence deserves credence 1. Yet, the fact that "P is a pointless truth" (for some P) can plausibly be part of one's evidence. If this is right, then agents could (often) assign credence 1 in "P is a pointless truth." This, however, would be impossible if rational agents could not believe with credence 1 that a proposition is pointless.

Moreover, our opponent is forced to explain away the many intuitive cases in which it seems that agents can be rationally certain that a proposition is pointless.

³³ This is so, because $E(W|q)$ can be conditionalized on $W=0$. Formally: $p(q) \cdot E(W|q, W=0)=0$. See footnotes 23–25 and 27.

³⁴ This follows from the Simple Principle of Conditionalization (Talbot, 2016, secs. 2, 6.2.A). This also follows from any view which says that agents can or should assign credence 0 to propositions that are inconsistent with their evidence. See, e.g., Williamson (2000, chap. 9) and Titelbaum (2013, 84).

Think, for instance, of the case we discussed above. Suppose you know for sure that you will die in the next 10 s. Given these extreme circumstances, it seems entirely permissible to have credence 1 in “Learning the number of giant sequoias in Muir Woods in the next 10 s is practically pointless for me.” If rational agents can have credence 1 in anything, then it is hard to see why this would not include propositions like this.

All we need for our argument to succeed is that, *in some cases*, rational agents can have credence 1 in “P is a pointless truth” (for some P). The burden of proof falls on the objector to show that it is *impossible* to have a rational credence 1 in “P is a pointless truth” (for some P).

5 Conclusion

Epistemic consequentialists are typically committed to two monistic theses. The first is a veritistic or accuracy-based form of monism about epistemic value: there is not a plurality of final epistemic values besides accuracy. The second is a form of monism about epistemic scoring: there is not a plurality of legitimate systems of epistemic scoring besides strictly proper scoring rules (SPSR).

We have argued that there is a tension between those two central tenets of VEC. This tension is revealed once we take seriously the following questions: why view SPSR as the only legitimate system to measure inaccuracy? In general, the scoring system(s) we should choose is determined by the objective(s) we want to reward. For veritists, the sole epistemic objective to be rewarded is accuracy. Hence, they ought to choose epistemic scoring systems that reward accuracy. But while SPSR do reward accuracy, some non-SPSR also appear to reward it. So, why should veritists exclude these non-SPSR? We have examined four prominent answers from proponents of VEC and found all of them wanting. Therefore, unless epistemic consequentialists come up with a vindication of SPSR that really does trace back to the need to reward accuracy, then they will have to give up at least one of their two central monistic commitments. Perhaps SPSR really is the only legitimate epistemic scoring system, but because mere accuracy is not, in fact, the sole final epistemic value to be rewarded. Or perhaps accuracy really is the sole final epistemic value, but non-SPSR sometimes are legitimate ways of measuring inaccuracy. In any case, this remains an unresolved tension at the core of the veristic epistemic consequentialist research programme.

Acknowledgements The authors received financial support from the Social Sciences and Humanities Research Council (Grants #756-2019-0133 and #756-2019-0366).

References

- Ahlstrom-Vij, K., & Dunn, J. (2018). *Epistemic consequentialism*. Oxford University Press.
- Berker, S. (2013). Epistemic teleology and the separateness of propositions. *Philosophical Review*, 122(3), 337–393.

- Bickel, J. E. (2007). Some comparisons among quadratic, spherical, and logarithmic scoring rules. *Decision Analysis*, 4(2), 49–65.
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1), 1–3.
- Broome, J. (2013). *Rationality through reasoning*. Wiley.
- Budescu, D., & Bar-Hillel, M. (1993). To guess or not to guess: A decision-theoretic view of formula scoring. *Journal of Educational Measurement*, 30(4), 277–291.
- Christensen, D. (2013). Epistemic modesty defended. In D. Christensen & J. Lackey (Eds.), *The epistemology of disagreement: New essays*. (pp. 77). Oxford University Press.
- Coates, A. (2012). Rational epistemic akrasia. *American Philosophical Quarterly*, 48(2), 113–124.
- Côté-Bouchard, C. (2017). Is epistemic normativity value-based? *Dialogue*, 56(3), 407–430.
- D'Agostino, M., & Dardadoni, V. (2009). What's so special about Euclidean distance? *Social Choice and Welfare*, 33(2), 211–233.
- D'Agostino, M., & Sinigaglia, C., et al. (2010). Epistemic accuracy and subjective probability. In M. Suárez (Ed.), *EPSA Epistemology and methodology of science* (pp. 95–105). Springer. https://doi.org/10.1007/978-90-481-3263-8_8
- Daoust, M.-K. (2020). Should agents be immodest? *Analytic Philosophy*. <https://doi.org/10.1111/phib.12194>
- David, M. (2005). On 'truth is good.' *Philosophical Books*, 46(4), 296.
- DiPaolo, J. (2018). Second best epistemology: Fallibility and normativity. *Philosophical Studies*, 176(8), 2043–2066. <https://doi.org/10.1007/s11098-018-1110-y>.
- Dorst, K. (2017). Lockean maximize expected accuracy. *Mind*, 128(509), 175–211. <https://doi.org/10.1093/mind/fzx028>.
- Dunn, J. 2015. Epistemic consequentialism. In *Internet Encyclopedia of Philosophy*. <http://www.iep.utm.edu/epis-con>.
- Easwaran, K. 2015. Dr. Truthlove or: How I Learned to Stop Worrying and Love Bayesian Probabilities. *Noûs*, 50(4), 816–853. <https://doi.org/10.1111/nous.12099/pdf>.
- Easwaran, K., & Fitelson, B. (2015). Accuracy, coherence, and evidence. *Oxford Studies in Epistemology*, 5, 61–96.
- Elga, A. (2010). How to disagree about how to disagree. In W. Ted & F. Richard (Eds.), *Disagreement*. Oxford University Press.
- Gibbard, A. (2007). Rational credence and the value of truth. *Oxford Studies in Epistemology*, 2, 143–164.
- Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378.
- Greaves, H. (2013). Epistemic decision theory. *Mind*, 122(488), 915–952.
- Greaves, H., & Wallace, D. (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind*, 115(459), 607–632.
- Hájek, A. (2009). Arguments for—or against—probabilism? In F. Huber & C. Schmidt-Petri (Eds.), *Degrees of belief* (pp. 229–251). Springer.
- Hawthorne, J., & Bovens, L. (1999). The preface, the lottery, and the logic of belief. *Mind*, 108(430), 241–264.
- Hazlett, A. (2013). *A luxury of the understanding: On the value of true belief*. Oxford University Press.
- Hedden, B. (2015). A defense of objectivism about evidential support. *Canadian Journal of Philosophy*, 45(5–6), 716–743.
- Horowitz, S. (2014). Immoderately rational. *Philosophical Studies*, 167(1), 41–56.
- Joyce, J. M. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science*, 65(4), 575–603.
- Joyce, J. M. (2009). Accuracy and coherence: Prospects for an alethic epistemology of partial belief. In F. Huber & C. Schmidt-Petri (Eds.), *Degrees of belief* (pp. 263–297). Springer.
- Kiesewetter, B. (2017). *The normativity of rationality*. Oxford University Press.
- Konek, J., & Levinstein, B. (2017). The foundations of epistemic decision theory. *Mind*, 128(509), 69–107. <https://doi.org/10.1093/mind/fzw044>
- Kvanvig, J. (2008). Pointless truth. *Midwest Studies in Philosophy*, 32(1), 199–212.
- Lasonen-Aarnio, M. (2018). Enkrasia or evidentialism? Learning to love mismatch. 177, 597–632. *Philosophical Studies*. <https://doi.org/10.1007/s11098-018-1196-2>
- Lesage, E., Valcke, M., & Sabbe, E. (2013). Scoring methods for multiple choice assessment in higher education. *Studies in Educational Evaluation*, 39(3), 188–193.
- Levinstein, B. (2017a). A pragmatist's guide to epistemic utility. *Philosophy of Science*, 84(4), 613–638.

- Levinstein, B. (2017b). Accuracy uncomposed: Against calibrationism. *Episteme*, 14(1), 59–69.
- Lewis, D. (1971). Immodest inductive methods. *Philosophy of Science*, 38(1), 54–63.
- Li, H. (2016). The trouble with having standards. *Philosophical Studies*, 176, 1225–1245. <https://doi.org/10.1007/s11098-018-1055-1>.
- Littlejohn, C. (2012). *Justification and the truth-connection*. Cambridge University Press.
- Littlejohn, C. (2015). Who cares what you accurately believe? *Philosophical Perspectives*, 29(1), 217–248. <https://doi.org/10.1111/phpe.12064>
- Lord, E. (2018). *The importance of being rational*. Oxford University Press.
- Lynch, M. (2004). *True to life: Why truth matters*. MIT Press.
- Mayo-Wilson, C., & Wheeler, G. (2016). Scoring imprecise credences: A mildly immodest proposal. *Philosophy and Phenomenological Research*, 92(1), 55–78. <https://doi.org/10.1111/phpr.12256/full>.
- Nelkin, D. K. (2000). The lottery paradox, knowledge, and rationality. *Philosophical Review*, pp. 373–409.
- Pettigrew, R. (2016). *Accuracy and the laws of credence*. Oxford University Press.
- Pettigrew, R. (2017). Précis and replies to contributors for book symposium on accuracy and the laws of credence. *Episteme*, 14(1), 1–30.
- Ramsey, F. (1931). Truth and probability. *The Foundations of Mathematics and Other Logical Essays*, pp. 156–198.
- Schechter, J. (2013). Rational self-doubt and the failure of closure. *Philosophical Studies*, 163(2), 429–452.
- Schervish, M. (1989). A general method for comparing probability assessors. *Annals of Statistics*, 17(4), 1856–79.
- Schoenfield, M. (2014). Permission to believe: Why permissivism is true and what it tells us about irrelevant influences on belief. *Noûs*, 48(2), 193–218.
- Talbott, W. (2016). Bayesian epistemology. In: Zalta E. N. (Ed.), *The Stanford encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/win2016/entries/epistemology-bayesian/>.
- Titelbaum, M. G. (2013). *Quitting certainties: A Bayesian framework modeling degrees of belief*. Oxford University Press.
- Wedgwood, R. (2017). *The value of rationality*. Oxford University Press.
- Whiting, D. (2013). The good and the true (or the bad and the false). *Philosophy*, 88(2), 219–242.
- Worsnip, A. (2018). The conflict of evidence and coherence. *Philosophy and Phenomenological Research*, 46(1), 3–44. <https://doi.org/10.1111/phpr.12246>
- Wrenn, C. B. (2017). Truth is not (very) intrinsically valuable. *Pacific Philosophical Quarterly*, 98(1), 108–128.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.