

# PERMISSIVISM, THE VALUE OF RATIONALITY, AND A CONVERGENCE-THEORETIC EPISTEMOLOGY

Ru Ye  
Wuhan University

Epistemic permissivism says that sometimes there are multiple rational responses to the same body of evidence. A recent argument against permissivism says that this view is incompatible with a plausible understanding of the accuracy-conduciveness of rationality, according to which rationality is accuracy-conducive because rational credence is more expectedly accurate than irrational credence. This is called ‘the value problem for permissivism.’ In this paper, I propose a new response to this problem. I defend a convergence-theoretic epistemology: Rationality is accuracy-conducive not because rational credence is more expectedly accurate than irrational credence, but because rational credence performs better with regard to convergence to truth. Drawing on recent developments in formal learning theory, I argue that this ‘convergence-to-truth’ understanding of the accuracy-conduciveness of rationality has many attractive features, and I argue that the convergence understanding is compatible with and even favors permissivism.

*Keywords:* Epistemic Permissivism, Epistemic Rationality, Truth-Conduciveness, Formal Learning Theory

## 1. Introduction

Can there sometimes be multiple rational responses to the same body of evidence? An impermissivist says no, while a permissivist says yes. The former thinks that rationality is strict—it doesn’t allow any leeway in how we should respond to our evidence. The latter thinks that rationality is slack—people with the same evidence can rationally respond differently, due to differences in factors such as epistemic standards, background beliefs, practical stakes, ways of weighing theoretical virtues, etc.<sup>1</sup> If we assume a Bayesian setting in which rational responses to evidence take the form of conditionalization, then the impermissivist thinks that there is a unique rational prior probability, whereas the permissivist thinks that there are many rational priors.

Most Bayesians are permissivists. In particular, subjective Bayesianism (which says that there are many rational priors) represents the position that’s called ‘interpersonal permissivism,’ since it implies that rational credence given a set of evidence is unique for a person but could vary across persons (Kelly 2014; Schoenfield 2014). However, recently,

---

<sup>1</sup> Recent defenders of impermissivism include Dogramaci and Horowitz (2016), Greco and Hedden (2016), Horowitz (2014, 2019), Schultheis (2018), and White (2014). Recent defenders of permissivism include Kelly (2014), Meacham (2019), Palmira (forthcoming), Schoenfield (2014, 2019), Thorstad (2019), and Ye (2019). Weisberg (2020) defends a mixed position: Evidential support is unique but rationality is permissive.

subjective Bayesianism has faced an important objection that is forcefully presented by Horowitz (2014, 2019). Roughly, the objection is that we think rationality is valuable because it's accuracy-conducive, and yet we can only make sense of the accuracy-conduciveness of rationality if we assume rationality is impermissive. This is the so-called 'the value problem for permissivism.'<sup>2</sup>

This paper proposes a new response to the value problem. Here is the plan. First, I explain the value problem in greater detail (Section 2). Then I propose a new response by defending a novel approach to rationality that I call 'convergence-theoretic epistemology' (Section 3). Drawing on recent developments in formal learning theory, I argue that this convergence approach has several attractive features and that it is compatible with and even favors permissivism (Section 4). Then I answer a series of worries about the convergence-theoretic epistemology (Section 5). Section 6 concludes the paper. I hope that this paper provides a novel understanding of the value of permissive rationality. I also hope that it will initiate further explorations of the convergence-theoretic epistemology, about what it can do and what it can imply for other debates in epistemology.

## 2. The Value Problem for Permissivism

Epistemologists often claim that rationality is truth-conducive, or, in the case of credence, they claim that rationality is accuracy-conducive. It's not clear what exactly that means. One natural understanding is that rational credence functions are more accurate than irrational credence functions *on average*. And since what happens on average could be expressed as what happens *in expectation* relative to some probability function, the above understanding of the accuracy-conduciveness of rationality is often put as the claim that rational credence has greater expected accuracy relative to some relevant probability function  $p$  (Horowitz, 2014, 2019; Schoenfield, 2015; Wedgwood, 2017, Ch. 9).

What's the probability function  $p$  here? Most scholars in the debate on the value of permissive rationality (which include both the impermissivist Horowitz and the permissivist Schoenfield) happen to adopt some internalist stance: they have interpreted  $p$  as a *rational credence function*. On this interpretation, to say that rationality is accuracy-conducive is to say that, relative to a body of evidence, a rational credence has greater expected accuracy than an irrational credence *from any rational point of view*. (Note that here we are silent on whether the rational point of view is unique.) We can put this idea more precisely as follows.

### Expectation Constraint

---

<sup>2</sup> Note that the value problem discussed in this paper is about the value of *being* rational in the pursuit of truth. This is different from the value problem discussed by Dogramaci and Horowitz (2016), which is about the value of *attributing* rationality—they argue that permissivists cannot explain the value of our practice of attributing rationality or irrationality to one another in society, because such practice would not be efficient in promoting social coordination if permissivism were true. Similarly, the challenge for permissivism raised by Levinstein (2017) and Greco and Hedden (2016) also concerns rationality-attribution: They argue that permissivists can't explain the value of rationality attribution in planning and deference. For permissivists' responses to the value problem of rationality-attribution, see Thorstad (2019) and Meacham (2019).

Let  $E$  be an arbitrary body of evidence. For *any* credence function  $c$  that is rational given  $E$ , if  $c$  assigns 1 to ‘credence  $cr$  is rational given  $E$  and  $cr^*$  is irrational,’ then relative to  $c$ , the expected accuracy of  $cr$  is greater than that of  $cr^*$ .

A principle like the Expectation Constraint has been implicitly accepted by many scholars (Horowitz, 2014, 2019; Wedgwood, 2017, Ch. 9; Schoenfield, 2015, pp. 636–8). Horowitz (2014; 2019) argues that permissivism faces a value problem because it’s incompatible with this constraint: If you are in an acknowledged permissive case, you will regard some irrational credence as more expectedly accurate than some rational credence. Here is a reconstruction of her example.

Suppose that you are in an acknowledged permissive case about a proposition  $H$ ; in particular, suppose that you are certain that the range of rational credence in  $H$  is  $[0.6, 0.9]$ . Then, unless you happen to adopt the midpoint credence of this range, you will regard some credence outside the range more expectedly accurate than some credence inside the range. For instance, suppose your credence in  $H$  is 0.9. Then you would regard 0.95 more expectedly accurate than 0.6. This is because our accuracy measure has this feature: your expectation of how accurate a credence is depends on how close that credence is to your own credence—you would expect closer credence to be more accurate.<sup>3</sup> (In particular, our accuracy measure is generated by ‘strictly proper’ scoring rules so that a probabilistically coherent person would expect her own credence to be uniquely maximally accurate.) Due to this feature, if you don’t happen to occupy the midpoint of the rational range, there will be credence outside the range that’s closer to you than some credence within the range, and thus there will be irrational credence that you expect to be more accurate than some rational credence.

So, permissivism is incompatible with the Expectation Constraint. In contrast, impermissivism *is* compatible with the constraint.<sup>4</sup> If impermissivism were true, any rational agent would regard a rational credence as more expectedly accurate than an irrational credence, because any rational agent can make the argument ‘my own credence is rational; if  $cr$  is also rational then it matches my own credence given impermissivism; my own credence uniquely maximizes expected accuracy given strict propriety; thus, if  $cr$  is rational, it must uniquely maximize expected accuracy’ (Horowitz, 2014, pp. 46–7; 2019, pp. 256–61).

Permissivists have offered several responses to the value problem. Some target the above argument that impermissivism is compatible the Expectation Constraint. They claim that the argument falsely assumes that rational agents are always certain of their own rationality, or that it suffers vicious circularity, or that it falsely assumes that rational agents always have precise credence (Schoenfield, 2019, p. 290; Thorstad, 2019, p. 911). Others deny the Expectation Constraint. For instance, Schoenfield (2019, fn. 16) suggests a ‘bracketing strategy,’ which says that rational agents sometimes need to bracket their own credence in evaluating the expected accuracy of others’ credence. So, it’s not true that

---

<sup>3</sup> This point clearly holds relative to the accuracy measure Brier score and the squared Euclidean distance that generates it. For a more general argument for this point, see Pettigrew (2013, pp. 32–3). (His proof would still go through if we replace the chance function in it with an arbitrary probability function.)

<sup>4</sup> Some impermissivists (such as Horowitz (2019)) even claim that their view *explains* the constraint.

rational credence should maximize expected accuracy from *all* rational perspectives. Schoenfield (2019) also defends an ‘endorsement proposal’ as an alternative account of the accuracy-conduciveness of rationality and argues that this proposal favors permissivism.

Existing impermissivists have not been convinced by the above responses. For instance, Horowitz (2019) has raised several doubts about the bracketing strategy and the endorsement proposal. In this paper, I won’t delve into the issue of whether permissivists can overcome those doubts.<sup>5</sup> Rather, I will use the value problem as an opportunity to motivate a convergence-theoretic approach to rationality. I will argue that this new approach provides a promising understanding of the accuracy-conduciveness of rationality and that it’s compatible with and even favors permissivism.

### 3. A New Response: Convergence-to-Truth as Accuracy-Conduciveness

#### 3.1 An Initial Motivation For a Convergence-Theoretic Epistemology

Suppose that two of your friends, Hanna and Leo, share evidence with you. Currently, Hanna’s credence function is  $cr_1$ , Leo’s is  $cr_2$ , and from your current point of view,  $cr_1$  is less expectedly accurate than  $cr_2$ . Can there still be a plausible sense of accuracy-conduciveness in which you regard Hanna’s credence as more accuracy-conducive than Leo’s?

Here is a reason for a positive answer. Even if you don’t regard Hanna’s credence  $cr_1$  as more expectedly accurate than Leo’s credence  $cr_2$  *in your current evidential situation*, you might know that, by relying on her credence  $cr_1$  as priors, Hanna will update to more and more accurate credence as *more evidence is gained*, but Leo won’t update to more and more accurate credence by relying on his credence  $cr_2$  as priors. That is, you might know that Hanna will *converge to truth* if she updates on an increasing body of evidence whereas Leo won’t converge to truth given the same increasing body of evidence. If so, then even if you regard  $cr_1$  as less expectedly accurate than  $cr_2$  in the *current* evidential situation, you can still regard  $cr_1$  as a better ‘learning tool’ than  $cr_2$  in pursuing accuracy; so, in some sense of accuracy-conduciveness, you can regard  $cr_1$  as more accuracy-conducive.

I will explain this convergence understanding of accuracy-conduciveness more precisely below, but here is a crude example to explain how it differs from the Expectation Constraint. Suppose that your current credence in a proposition  $p$  is 0.3, Hanna’s credence in  $p$  is 0.9, and Leo’s credence in  $p$  is 0. Then you regard Hanna’s credence in  $p$  as less expectedly accurate than Leo’s credence. However, assuming that Hanna meets certain conditions (which are explained below), you can be certain that, if  $p$  were true, her credence in  $p$  will approach certainty as she gains more evidence, and if  $p$  were false, her credence in  $p$  will approach 0 as she gains more evidence. By contrast, you are certain that Leo won’t perform so well as more evidence is gained—since his current credence in  $p$  is 0, it will remain 0 no matter how much more evidence he gains. So, you are certain that Hanna will converge to the truth about  $p$  in both the  $p$ -worlds and the not- $p$  worlds, but Leo won’t converge to the truth in the  $p$ -worlds. In this case, it seems that there is an important sense in which you can regard Hanna’s credence as more accuracy-conducive than Leo’s credence, even if you regard it as less expectedly accurate in your current evidential situation.

---

<sup>5</sup> See Palmira (forthcoming) for an attempt to overcome Horowitz’s (2019) doubts about the endorsement proposal.

This example suggests an understanding of the accuracy-conduciveness of rationality different from the Expectation Constraint. According to this new understanding, the value of rationality in pursuing accuracy doesn't lie in how expectedly accurate a rational credence is in a static evidential situation, but in how rational persons perform in a dynamic situation where they get more and more evidence. Rational persons, even if they are highly inaccurate in a static evidential situation, will *become* more and more accurate by updating on an increasing body of evidence, whereas there is no such guarantee for irrational persons. I call this understanding of the accuracy-conduciveness of rationality 'Convergence Constraint,' which says,

Convergence Constraint (roughly formulated)

A rational credence function performs better than an irrational credence function with respect to convergence to truth.

In what follows, I will do three things. First, I clarify the notion of 'convergence to truth' and I explain what counts as 'performing better' with respect to convergence to truth. Second, I argue that the Convergence Constraint is an attractive view about the accuracy-conduciveness of rationality. Third, I argue that the Convergence Constraint is compatible with and even favors permissivism.

### 3.2 Clarifying the Convergence Constraint

Before I get started, I should note that, although my primary aim in this paper is to defend permissivism, I also aim to urge scholars in the accuracy-centered epistemology to pay more attention to the convergence understanding of accuracy-conduciveness. Currently, much of the work in the accuracy-centered epistemology focuses on justifying rationality norms by showing that they are accuracy-conducive in some sense, using mostly decision-theoretic norms like non-dominance, maximizing expected accuracy, or maxi-min (Joyce, 1998; Greaves & Wallace, 2006; Pettigrew, 2016). Norms involving convergence-to-truth haven't yet received much attention. But convergence to truth is an interesting and fruitful understanding of the elusive notion of truth-conduciveness—this is the reason why the convergence approach has become a major approach in formal learning theories, theories designed to study what kind of learning methods are most reliable and efficient in satisfying our cognitive goals. So, it's worth exploring the ramifications of accepting convergence to truth as an understanding of the truth-conduciveness of rationality.<sup>6</sup>

Now, what is 'converging to truth' exactly? I will clarify this notion using a typical model in Bayesian versions of formal learning theory. The first basic concept is a 'learning problem.' A learning problem can be identified with a pair  $(W, F)$ , where  $W$  is a set of possibilities compatible with one's current evidence and  $F$  is a set of hypotheses to be

---

<sup>6</sup> An early application of formal learning theory to traditional epistemological problems is Putnam's (1963) criticism of Carnap's measure of confirmation, where Putnam argues that Carnap's measure is inadequate because it doesn't satisfy the condition that the 'instance confirmation' of a true hypothesis converges to 1 (or even stays above 0.5) as more and more observations are made. For contemporary philosophy-oriented introductions of formal learning theory, see Juhl (1997), Kelly et al. (1997), and Schulte (1999a, 1999b). For introductions that include more formal details, see Kelly (1996).

learned. Each possible world in  $W$  specifies a possible infinite stream of evidence, a way of how the world might unfold for one from now on as time grows indefinitely. And a hypothesis in  $F$  is a subset of  $W$ . The second basic concept is a ‘learning method.’ One’s learning method is identified with a function from *initial segments* of evidence streams in  $W$  to credence functions. This is meant to capture the fact that, at any finite stage of our inquiry, we can’t observe an infinite evidence stream in its entirety but only its initial, finite segment. A Bayesian agent’s learning method is encoded in her current credence function. In each world  $w$ , her credence function updates sequentially by conditionalizing on the initial segments of the evidence stream in  $w$ .

To illustrate, consider this toy model. Suppose that a Bayesian agent, whose current credence is represented by a probability function  $P$ , wants to know whether a coin is fair and he tries to figure it out by flipping the coin repeatedly, and suppose that his flipping could carry out indefinitely long. We can model this learning problem as  $(W, F)$ , where each point  $w \in W$  is a countably infinite binary sequence representing a possible outcome of the infinite coin toss such as (HHTH...),  $F$  is a set of subsets of  $W$  that represents propositions the agent is interested in learning. Note that the learning problem must be compatible with the agent’s current evidence, in the sense that  $W$  only includes worlds that are consistent with her current evidence. At each later time  $n$ , the agent tosses the coin, notes the result, and thus expands his evidence. Let  $E_{nw} \subseteq W$  be the set of all binary sequences whose  $n$ -initial segment is the same as  $w$ ’s, so that it represents the agent’s evidence in  $w$  after  $n$  tosses. Finally, for each proposition  $A$  in  $F$ , let  $I_A(w)$  be the indicator function of  $A$  in  $w$ , that is,  $I_A(w) = 1$  if  $A$  is true in  $w$  and  $I_A(w) = 0$  otherwise. To say that the Bayesian agent converges to the truth in world  $w$  as he gets more evidence is to say that, for any  $A$  in  $F$ ,

$$\lim_{n \rightarrow \infty} P(A | E_{nw}) = I_A(w)$$

That is, an agent converges to truth in  $w$  if, as the agent gets more evidence in  $w$ , his posteriors function approaches the omniscient credence function in  $w$ .

With the notion ‘convergence to truth in a world’ explained, we can explain what it is for a credence to ‘perform better’ with regard to convergence to truth than another credence. We do this by distinguishing various modes of convergence:

- a. Speed of convergence: convergence can happen fast or slow.
- b. Domain of convergence: convergence can happen on a larger or smaller set of worlds in the state space.
- c. Stability of convergence: convergence can be stable or unstable.<sup>7</sup>
- d. Frequency of retraction in convergence: you might need to change your mind more (or less) frequently in converging to truth.<sup>8</sup>

---

<sup>7</sup> A credence function converges to truth stably in a world  $w$  if, at the *first* time it updates to a high confidence in the truth in  $w$ , it continues to update to high confidence in the truth in  $w$ . That is, once it outputs the truth it will never let the truth go.

<sup>8</sup> A credence function ‘changes its mind’ on a proposition  $A$  at some stage  $n$  if it’s highly confident in  $A$  at time  $n-1$  and yet highly confident in not- $A$  at  $n$ .

Among these modes of convergence, some are clearly desirable while others are not so clearly desirable. For instance, speed-optimal convergence, convergence in a maximal domain, and convergence that minimizes possible times of retraction all seem desirable. But it's not clear whether stability in convergence is desirable. This issue will not be settled here, since my later arguments don't hang on it. The important point is that, using the above distinction among various modes of convergence, we can clarify 'performing better in convergence' and clarify the Convergence Constraint as follows:

Convergence Constraint (better formulated)

A rational credence function given a set of evidence performs better than an irrational credence function with regard to convergence to truth in the following sense: in solving those learning problems that are compatible with the evidence, a rational credence achieves all the desirable modes of convergence that can be achieved whereas an irrational credence doesn't; or, when it's impossible to achieve all the desirable modes of convergence in a learning problem, a rational credence strikes a better balance with regard to those modes than an irrational credence does.

Let's call a theory of rationality that makes the above constraint its essential commitment 'a convergence-theoretic epistemology.' In the following section, I argue that this novel epistemology has several attractive features. Before that, I will address two immediate worries.

First, you might worry that the constraint would make rationality relative to learning problems, since the convergence properties of a learning method are characterized relative to a learning problem modelled as  $(W, F)$ , and it's possible that no learning method does better than others for all learning problems.

However, this consequence of relativity is acceptable. Note that, on the above model of learning problems, the goodness of a method is relative to what evidential streams are possible (i.e., what  $W$  includes) and relative to what hypotheses you are interested in learning (i.e., what  $F$  includes). The first aspect of the relativity is perfectly acceptable—since elements of  $W$  are fixed by one's evidence, this just means that what credence functions are rational is relative to one's evidence. The relativity to what we are interested in learning is also acceptable. If adopting a credence function at least partly means adopting a learning method, then just like whether any old method is good is relative to what problems it's used to solve, whether a learning method is good is relative to what learning problems it's used to solve. Consider an extreme example. If, for some reason, you are just not interested in learning any empirical hypothesis, then presumably it doesn't matter whether you are an inductivist or a counterinductivist.

Moreover, even if you don't like the above kind of relativity, we can revise the Convergence Constraint to avoid it, by adding a criterion of 'no weak-dominance': Given a body of evidence, a rational method  $M$  is not weakly dominated by other methods with respect to the hypotheses to be learned; that is, no other method does no worse than  $M$  for all the hypotheses to be learned and does better than  $M$  for some hypotheses.

There is a second immediate worry: Perhaps the Convergence Constraint is not a substantive constraint on rationality at all, because of the Bayesian ‘washing out’ theorems. Basically, these theorems say that differences in Bayesian agents’ priors don’t matter because eventually they all converge to truth under certain reasonably weak conditions.<sup>9</sup>

This worry is misguided. Bayesian washing out theorems all concern ‘almost sure’ convergence; that is, they all have the form that, under certain conditions, any Bayesian agent assigns 1 to the proposition that she will converge to truth. So they only say that Bayesian agents will *think* that they will converge to truth, not that they will converge to truth *in fact*, and the kind of convergence that formal learning theorists care about is convergence to truth in fact.<sup>10</sup> And convergence from one’s own point of view can’t give rise to the kind of objectivity or truth-conduciveness we want: if one’s credence function merely allows one to assign 1 to the proposition that one will converge to truth while in fact it doesn’t converge to truth in a large region in the state space, one’s credence function doesn’t seem to be genuinely truth-conducive (Earman, 1992, pp. 147–8; Kelly et al., 1997, pp. 261–3).<sup>11</sup>

### 3.3 Interesting/Attractive Features of the Convergence-Theoretic Epistemology

The idea that we could evaluate learning methods (also known as ‘inductive inferential rules’) by studying their convergence properties has a lot of supporters outside of mainstream epistemology. Besides formal learning theorists, those doing large sample theory in statistics also study convergence properties: In estimator-evaluation, they focus on convergence properties such as statistical consistency, unbiasedness, efficiency, etc. In what follows, I list some interesting features of convergence-theoretic epistemology. I admit that the attractiveness of some of those features might be more controversial than that of the others, but I hope to convince you that the features viewed as a package make convergence-theoretic epistemology an interesting theory that deserves further exploration.

**3.3.1 Explaining the value of rationality dynamically.** Convergence-theoretic epistemology shares a core commitment of the current decision-theoretic accuracy-centered epistemology, which says that one’s credence is rational only if it’s a good tool for achieving one’s accuracy goals. But when it comes to what makes a credence function a good tool for achieving accuracy goals, convergence-theoretic epistemology has a different focus than the current decision-theoretic accuracy-centered epistemology: rather than focusing on the

---

<sup>9</sup> For instance, Savage (1972, pp. 46–50) proves that, in those experiments where observations have independent and identical distributions, any Bayesian agent converges to truth almost surely. Hawthorne (2018) strengthens Savage’s proof by relaxing the ‘identical distribution’ condition. Gaifman and Snir (1982) prove that, if the evidence one is about to accumulate ‘separates’ the state space, then one almost surely converges to truth.

<sup>10</sup> The most important mode of convergence for formal learning theorists concerns domain of convergence: A learning method does well with regard to this mode if it achieves ‘sure convergence’ (or what is called ‘logical reliability’), which takes the form:

$$\text{For every } w \text{ in } W, \lim_{n \rightarrow \infty} P(A | E_{nw}) = I_A(w)$$

This differs sharply from ‘almost sure convergence,’ which takes the form:

$$P(\{w \in W : \lim_{n \rightarrow \infty} P(A | E_{nw}) = I_A(w)\}) = 1.$$

<sup>11</sup> It’s worth noting that emphasizing actual convergence is compatible with internalism about rationality: As we will see in the next subsection, it’s knowable a priori whether a learning method has better convergence performance than another learning method.

benefit of rationality in a static evidential situation, it focuses on the benefit of rationality in a dynamic situation where evidence grows. This focus on dynamicity is inspired by an ordinary observation: when a person employs a rational method and yet fails to find the truth, we sometimes attribute her failure to her lacking sufficient evidence; that is, we think that if she gains more evidence she probably would find the truth. This suggests that *one* way of testing whether a learning method is good is to look at how well the method performs in a dynamic process where one's evidential situation evolves with time. Rational persons might not do well at initial stages of inquiry, but as evidence grows they will do better than irrational persons.

Of course, the convergence-theoretic epistemology focuses on an idealized version of the above dynamic picture: It assumes that the evidence could grow without bounds. Some readers might balk at this idealization. But as I will argue in Section 5, considering its usefulness, the idealization is no more obnoxious than the other kinds of idealizations that we are used to in doing Bayesian epistemology.

**3.3.2 Continuity with the Platonic value on stability.** Recall the basic idea behind the notion 'convergence-to-truth in a world.' The limit notation says that, for any positive distance from the truth, no matter how small, there is a finite time at which the method outputs credence within that distance, and it *continues to do so at all later times*. Putting the convergence idea in terms of binary beliefs, a method converges to truths in a world if it outputs true beliefs at some finite time and *continues to do so* at all later times; that is, not only it finds the truths at some finite time, it also settles on, or 'tethers' the truth from that time on. This idea of settling on the truth is not entirely new. It resonates with an old, Platonic stability theory of knowledge. In *Meno*, Plato argues that knowledge is more valuable than a mere true belief:

For these [true opinions], so long as they stay with us, are a fine possession, and effect all that is good; but they do not care to stay for long, and run away out of the human soul, and thus are of no great value until one makes them fast with causal reasoning ... But once they are fastened, in the first place they turn into knowledge, and in the second, are abiding. And this is why knowledge is more prized than right opinion, the one transcends the other by its trammels. (Plato, 1924, p. 363)

The idea here is that the stability of a true belief makes it more valuable than a mere true belief. And although Plato is talking about knowledge in the quotation, there is no reason why we couldn't extend the stability idea to rationality. Of course, there are many understandings of the stability. In epistemology, both the indefeasibility theory and the tracking theory of knowledge could be understood as ways of cashing out the stability idea—both emphasize that, to have knowledge, one must not only have a true belief in one's actual circumstance but also retain the true belief across a range of relevant circumstances. In researches about dynamic systems, the stability of a dynamic system sometimes refers to the system's ability to converge to equilibrium in the long run, which is the essential idea behind the traditional concept 'asymptotic stability' in stability analysis. Applying this idea of 'convergence to equilibrium' to learning methods, we get convergence-theoretic epistemology. The basic idea is that we can understand a person's belief-revision process as

an evolving, dynamic system guided by the person's learning method, and we can understand true beliefs as the equilibrium that the system strives for. Then the learning method is said to be good if, under its guidance, the dynamic system stabilizes on true beliefs after some finite time. So, convergence-theoretic epistemology could be viewed as a modern refinement of the old Platonic value of stability. (Note that the stability discussed here, which is achieved in every mode of convergence, should be distinguished from the 'stability mode of convergence' mentioned in Section 3.2.)

**3.3.3 Fruitfulness in recovering norms of rationality.** An important reason why current decision-theoretic, accuracy-centered epistemology is attractive lies in its success in recovering norms of rationality. As it happens, convergence-theoretic epistemology is also promising in recovering norms of rationality, although it still remains to be seen whether it can achieve the same scope of success as the current decision-theoretic epistemology does. For example, Schulte (1999b) argues that, in tackling Occam's problems (i.e., problems that have the structure of learning propositions that hypothesize the existence of something, such as certain particles or nonblack ravens), a learning method must satisfy Occam's Razor in order to achieve fast convergence on a maximal domain. He also argues that in learning problems discussed in the Goodman's Riddle of Induction, a learning method must do natural projection rather than grue projection in order to minimize retraction (i.e., minimizing the greatest possible times of retraction) in converging to truth. Similar arguments for simplicity are developed by Kelly in a series of papers (2007, 2010). More recently, Lin (forthcoming) argues that, in tackling learning problems that have a similar structure with the hard raven problem ('hard' because the candidate hypotheses include a Cartesian scenario in which there are nonblack ravens but they will never be observed), a learning method must implement enumerative induction rather than counterinduction or the skeptic's policy, in order to achieve stable convergence on a maximal domain. And Nielsen and Wofsey (2020) argue that, in learning problems concerning discovering the limiting relative frequency of an event, a credence function must satisfy a version of the reflection principle in order to achieve speed-optimal convergence.

To get an idea of how such vindication might go, let's see an example that's familiar from learning theorists' attempt to vindicate enumerative induction. Suppose that two persons, one enumerative inductivist and the other a counterinductivist, are interested in finding out whether all ravens are black, and they do so by sequentially observing an infinite sequence of ravens. Suppose that, before the observation, their priors have these properties: If they observe a nonblack raven, both will be certain that not all ravens are black; however, if all the  $n$  observed ravens at time  $n$  are black, the inductivist's credence in 'all ravens are black' will approach certainty as  $n$  approaches infinity (let's say that his credence at  $n$  will be  $1 - 1/(n+2)$ ), whereas the counterinductivist's credence in this proposition will never be greater than  $1/3$  (let's say that his credence at  $n$  will be  $1/3 - 1/(n+4)$ .) So, no matter how many ravens are uniformly observed to be black, the counterinductivist will always be more confident that not all ravens are black than that all ravens are black. Then the counterinductivist will perform worse with regard to convergence to truth than the inductivist: both converge to truth in worlds with nonblack ravens, but only the inductivist converges to truth in worlds with all black ravens—in such worlds they will observe

uniformly black ravens, and given such observation, the inductivist's credence in the truth will approach certainty, whereas the counterinductivist's credence in the truth won't be greater than 1/3. (Note that this fact about the two persons' convergence performance can be known a priori; that is, the counterinductivist can know *before the observation* that his own credence is worse than the inductivist in convergence to truth.) Since the counterinductivist's learning method doesn't achieve all the desirable modes of convergence that can be achieved, the Convergence Constraint implies that it's irrational.

**3.3.4 Perspective-neutrality in explaining the value of rationality.** A striking feature of convergence-theoretic epistemology lies in its perspective-neutrality in explaining the value of rationality. Recall that the Expectation Constraint discussed in Section 2 essentially says that a rational credence has greater expected accuracy than an irrational credence from the perspective of any *rational* person. So, it can't explain the value of rationality from the perspective of those *irrational* persons who also care about accuracy; thus, it can't explain why even irrational persons should also want to be rational. And this problem of perspective-relativity is common to any expected-accuracy explanation of the value of rationality, since expected-accuracy is always relative to some probability function.

The convergence-theoretic epistemology doesn't have this problem, because the Convergence Constraint can explain the value of rationality from the perspective of *any* coherent credence function. To see this, first note that whether a credence function has better convergence performance than another in a learning problem is a mathematical truth, not an empirical one—we only need to employ our knowledge of mathematical facts in reasoning about whether a credence function achieves convergence in a maximal domain, whether its convergence is speed-optimal, or whether it minimizes retraction. This differs sharply from our reasoning about expected accuracy—to know the expected accuracy of the counterinductivist's prediction 'the next emerald is not green,' we need to rely on our own opinion about the empirical matter 'the next emerald is not green.' Therefore, we cannot give an expected-accuracy argument against counterinduction without relying on our own inductive methods. In contrast, when a learning problem is given, reasoning about convergence performance doesn't rely on our views on empirical matters, and thus the reasoning is available for any probabilistically coherent person—including an irrational person like the counterinductivist. For example, in the above sub-section, the fact that the enumerative inductivist converges to truth in strictly more worlds than the counterinductivist does is a fact that even the counterinductivist herself can see, because argument for this fact has *not* employed enumerative induction—it's a deductive argument.

**3.3.5 Explaining why more evidence tends to be better.** It's well-known that accuracy-centered epistemologists can explain the value of evidence by invoking Good's theorem, which says that it's always rational to gather free evidence before acting. Convergence-theoretic epistemology can also explain why more evidence tends to be better for rational persons. As I've mentioned above, learning-theoretic epistemology has a dynamic character—what it demands on a rational method is not that it yields accurate beliefs at a static evidential situation but that it improves accuracy as evidence grows. Therefore, the rewards of being rational (i.e., the rewards of having good convergence properties) is

primarily reflected in the dynamic process of belief-revision as evidence grows. Then it's no wonder that more evidence tends to benefit rational persons.

To illustrate, consider an important desirable mode of convergence that learning theorists call 'logical reliability' (i.e., converging to truth in every possible world). If your learning method has this property, then you are guaranteed to approach the maximally accurate beliefs as you get more evidence, even if your method doesn't output highly accurate beliefs at the initial stages of inquiry. Moreover, even if your goal is not *maximally* accurate beliefs but merely beliefs whose accuracy passes certain threshold, more evidence still tends to be better, or at least not worse: for any desired level of accuracy, the finite stage of inquiry at which convergence to that level of accuracy begins is always a nondecreasing function of the desired level of accuracy.<sup>12</sup> So, it always takes no less evidence, and sometimes it takes more evidence, in order to stabilize on beliefs with a greater level of accuracy.<sup>13</sup>

(I should clarify that, to say that the benefit of rationality is reflected in a dynamic process of belief revision is not to say that you should extend your inquiry indefinitely long in order to reap that benefit. Since convergence, if achievable at all, always begins at some finite stage of inquiry, there is always a finite point at which you can stop your inquiry, because if you have converged to a desired level of accuracy then further inquiry won't make the accuracy of your belief fall below that level. So, more evidence is better only in the process of achieving convergence; once convergence has achieved and your belief has stabilized on the desired level of accuracy, further inquiry is not necessary. This accords with a familiar idea that knowledge, which is understood as a stable true belief in this context, marks the end of inquiry.)

In sum, convergence-theoretic epistemology has many interesting or attractive features: It provides a dynamic explanation of the value of rationality; it's continuous with the Platonic value on stability of true beliefs; it has many virtues that our current decision-theoretic accuracy-centered epistemology has, such as recovering norms of rationality and explaining why more evidence is better; in addition, it recovers norms of rationality in a perspective-neutral way. Of course, you might still find the defense offered above inadequate; in particular, you might worry that, by focusing on convergence properties, convergence-theoretic epistemology focuses too much on the long run and doesn't say much about the short run. This is an important concern, and I will return to it soon. But first let me

---

<sup>12</sup> To illustrate, for any  $w$  and any proposition  $H$ ,  $\min\{n: \text{for all } m \geq n, |P(H|E_{mw}) - I_H(w)| < 0.3\}$  could not be smaller than  $\min\{n: \text{for all } m \geq n, |P(H|E_{mw}) - I_H(w)| < 0.5\}$ . The proof is trivial:  $\min A \geq \min B$  if  $A \subseteq B$ .

<sup>13</sup> Some results in computational learning theory—a practical extension of formal learning theory that studies computational efficiency of learning methods—also tell us about the value of growing evidence. A fundamental theorem of computational learning theory tells us that, when your evidential stream is generated by some unknown objective probability function, the sample size that guarantees certain level of predictive accuracy at certain level of objective probability is an increasing function of the desired degree of accuracy and the desired objective probability. Therefore, if you want greater predictive accuracy or a fixed level of desired-accuracy at a greater probability no matter what the underlying objective probability function is, you will need more evidence. For a precise statement of the theorem that tells us the above result, see Kononenko & Kukar (2007, pp. 402–4).

explain why convergence-theoretic epistemology is compatible with and even favors permissivism.

#### 4. The Convergence Constraint Favors Permissivism

In section 2, we've seen that permissivism is incompatible with the Expectation Constraint, and we've seen that the incompatibility arises because, for some coherent credence within the rational range, a credence outside the range could be closer to it than some insider, and a coherent person always regards closer credence as more expectedly accurate. In particular, given strict propriety, a coherent person always regards her own credence as uniquely maximizing expected accuracy.

Similar problem wouldn't arise for Convergence Constraint, for a coherent person needs not view closer credence as having better convergence performance. Indeed, we've seen from the above vindication of enumerative induction that a counterinductivist can see that his convergence performance is worse than that of the enumerative inductivist, even though his own credence is clearly the closest to his own credence. In general, the vindication arguments for norms of rationality listed in the above section imply that a credence function's convergence performance depends on whether it satisfies the vindicated norms of rationality. So, even if credence function A is closer to you than credence function B is, if A doesn't satisfy some norms of rationality (such as reflection) but B satisfies all norms, then you could still view A as having worse convergence performance than B.

So, if the Convergence Constraint were incompatible with permissivism, it cannot be due to a reason that is similar to the reason why the Expectation Constraint is incompatible with permissivism. Are there any other reasons to think that they are incompatible? I can see none. Nothing in the constraint rules out the possibility that, in some learning problems, two different credence functions both have good convergence performance. After all, the convergence mode of a credence function is not always sensitive to small variance in the precise values the function assigns to each proposition. Consider, for instance, the mode of convergence about speed. Two different credence functions could converge to truth equally fast, according to the following natural way of defining speed of convergence.

Recall the learning problem modelled as  $(W, F)$  in Section 3.2. For any proposition A in F, any real number  $\epsilon$  in  $[0, 1]$  and any world  $w$  in  $W$ , we can define a function  $C_{\epsilon w A}(P)$ , which measures how fast one's current credence P has its posterior on A falling within  $\epsilon$ -range of the truth value of A in  $w$ , as the least number of experiments P needs in order for its posterior to fall within the range. That is:

$$C_{\epsilon w A}(P) = \min \{n: \text{for all } m \geq n, |P(A|E_{mw}) - I_A(w)| < \epsilon\}$$

Then we can consider another probability function Q defined on F, and we can define 'one converges to truth faster than the other' as:

P converges to the truth about A faster than Q if:

$$\begin{aligned} C_{\epsilon w A}(P) &\leq C_{\epsilon w A}(Q) \text{ for all } \epsilon \text{ in } [0, 1] \text{ and all } w \text{ in } W, \text{ and} \\ C_{\epsilon w A}(P) &< C_{\epsilon w A}(Q) \text{ for some } \epsilon \text{ in } [0, 1] \text{ and some } w \text{ in } W. \end{aligned}$$

Last, we can say that P and Q converge to the truth of A at an equal speed if neither one converges faster than the other.<sup>14</sup> It's easy to verify that, on this definition of speed of convergence, two different credence functions could converge to the truth of a target proposition at an equal speed. In fact, even for some designated  $\epsilon$  and some designated  $w$ , there could be two probability functions P and Q such that  $C_{\epsilon w A}(P) = C_{\epsilon w A}(Q)$ . For example, suppose  $\epsilon = 0.1$  and the designated world  $w$  is a world in which A is true. This situation is possible: On the 10<sup>th</sup> experiment, a Bayesian agent updates his credence from 0.5 to 0.9, and another Bayesian agent updates her credence from 0.6 to 0.9. That is, both agents take 10 experiments to make their posteriors on A fall within the range [0.9, 1].

In fact, the Convergence Constraint is not only compatible with permissivism but also favors it. For there are learning problems in which not all the desirable modes of convergence can be achieved, and thus a trade-off among those modes must be made. For example, Schulte (1999a, pp. 404–7) has pointed out that, in the simple learning problem that he calls ‘one-shot Riddle of Induction,’ the goal of minimizing retractions comes into conflict with the goal of fast convergence.<sup>15</sup> If we assume that there is no unique best way to trade off the desirable modes of convergence, then rational credence functions given a body of evidence will not be unique since credence functions with the best convergence performance in a learning problem would not be unique—some might converge to truth faster, while others might converge to truth with less retractions.

This argument is reminiscent of a familiar defense of permissivism that focuses on trade-offs among different epistemic values (e.g., seeking truths and avoiding errors) or among different theoretical virtues. As some permissivists have argued, relative to the same body of evidence, some credence functions are simpler while others are more explanatory; and since there doesn't seem to be a unique best way to trade off these two virtues, rational credence functions need not be unique (Titelbaum, 2015, p. 674). In a similar spirit, the argument in the above paragraph says that, given the same evidence, different credence functions could be equally rational because there are many ways of being truth-conducive

---

<sup>14</sup> The characterization of convergence speed of credence functions here is inspired by Schulte's (1999a, p. 413) definition of ‘data-minimal’ learning methods that concern binary beliefs. Nielsen and Wofsey (2020, p. 4) use a similar characterization.

<sup>15</sup> Here is the basic idea. In one-shot Riddle of Induction, there are only two competing hypotheses: all emeralds are green, or all emeralds are grue, where ‘grue’ is defined relative to a critical time  $t$ ; And we obtain conclusive evidence at time  $t$ : if the emerald observed at  $t$  is green, we know for certain that not all emeralds are grue; if the emerald observed at  $t$  is blue, we know for certain that not all emeralds are green. Consider two learning methods, a cautious one and a natural one. The cautious method says that we wait for the conclusive evidence at  $t$  before making a projection; the natural method says that we project the green hypothesis from the beginning and keep doing so until we obtain conflicting evidence, and if that happens we change mind and project the grue hypothesis. The cautious method minimizes retraction, since it doesn't have to retract no matter which hypothesis is true and the natural method might have to retract once; but the natural method dominates the cautious one in speed of convergence: if the grue hypothesis is true then both methods converge to truth at  $t$  and not before then; but if the green hypothesis is true, then the natural method converges to truth from the beginning and the cautious method doesn't converge before  $t$ .

(that is, there are many desirable modes of convergence to truth) and different credence functions could be truth-conducive in different ways.

## 5. Worries About the Convergence-Theoretic Epistemology

There is a worry lurking in the above argument that the Convergence Constraint is compatible with and even favors permissivism. The above argument might make you worry that learning-theoretic epistemology is too lenient, and thus it would lead to a form of permissivism that's too radical. After all, the Convergence Constraint seems to primarily focus on the long run and doesn't say much about the short run. That is, it focuses on what adopting a learning method implies *in an infinitely long sequence* of inquiry, where evidence grows *without bounds*, and it ignores the method's implication in much shorter run, such as when the evidence is fixed or grows with bounds.

Now, let me note that this worry about the long-run is not a single worry but consists of several separate, interrelated worries. First, the worry might be that learning-theoretic epistemology allows any behavior in the short-run as long as the long-run behavior is good. Or, the worry might be that learning-theoretic epistemology just couldn't be a correct epistemology, since long-run considerations shouldn't be relevant at all given the famous Keynesian warning 'in the long run, we are all dead.' Or, the worry might be that, even if long-run considerations are important, they just couldn't do enough explanatory work. In particular, they can at most explain the benefits of rationality in unbounded dynamic evidential situation and couldn't explain the benefits of rationality in static evidential situations. In what follows, I respond to the three worries in turn.

### 5.1 Anything Goes in the Short-Run?

The response to the first worry is easy: It's not the case that learning-theoretic epistemology allows anything to go in the short run. Whether a learning method converges to truth at all indeed depends only on how it behaves in the long run, but some desirable modes of convergence also depends on its behavior in the short run. To illustrate, consider the mode of convergence concerning speed, which has played an essential role in Nielsen and Wosfe's argument for the reflection principle. This mode clearly depends on the short-run. For two credence functions that both converge to the truth in a world, one might take no more than 100 experiments to update to a high credence in the truth of the world, whereas the other might take no more than 10 experiments. So, their behaviors in the first 100 experiments are clearly relevant to how fast they converge to the truth in a world.

Another mode of convergence that involves the short-run is 'minimaxing retraction,' that is, minimizing the greatest possible times of retraction. And this mode of convergence has played a significant role in learning theorists' argument for rationality norms like Occam's Razor (Kelly 2007, 2010; Schulte, 1999b). Here is an example originated from Schulte (1999b, pp. 17–18). Suppose that two persons face a learning problem in which the hypotheses are from {All emeralds are green; All emeralds are grue(1); All emeralds are grue(2), ...}, where an emerald is grue(n) if either it's green and observed before time n, or it's blue and not observed before n. Suppose that the first person projects naturally: she starts with high confidence in 'all emeralds are green' and continues to hold high confidence until a

blue emerald is observed at some time  $n$ ; if that happens she changes mind and switches to high confidence in the competing hypothesis ‘all emeralds are  $\text{grue}(n)$ .’ By contrast, the second person starts with high confidence in ‘all emeralds are  $\text{grue}(2)$ ,’ and if the first two observed emeralds are green, then he changes mind and switches to high confidence in ‘all emeralds are green’ and starts to behave like the first person since then. Then this person’s mode of convergence is not as good as the first person’s in avoiding retraction: the first person needs to change mind at most once whereas the second person might need to change mind twice—for instance, in the world in which emeralds are (Green, Green, Green, Blue, Blue, ...), the second person needs to change mind at both time 2 and time 4. So, although both persons converge to truth eventually, their short run performance is relevant in determining their mode of convergence concerning frequency of retractions.

## 5.2 How Could Long-Run Considerations Even Be Relevant?

Here is the second worry. Convergence considerations are all about what happens in the limit. It’s good that they imply some short-run constraints, but this is not good enough, for short-run considerations should be our *only* considerations. As Keynes has famously said, ‘in the long run, we are all dead.’ So, all considerations about what happens in the limit are irrelevant to rationality and thus convergence-theoretic epistemology must fail.

In reply, first note that caring about the long-run is not unique to convergence-theoretic epistemology; rather, it crops up in a wide variety of other epistemological theories. The widespread use of long-run defense gives us a reason to think that long-run considerations are not as illegitimate or irrelevant as they first seem.

As a first example, consider the norm of maximizing expected utility. Although this norm has been defended on multiple grounds such as representation theorems or the norm’s widespread applications, the long-run defense invoking the strong law of large numbers remains an important motivation (Briggs, 2019). It says that, if you maximize expected utility in a decision problem, then by the strong law of large numbers, you will almost surely maximize average actual utility if you stick with the same choice in an infinitely long repetition of the same decision problem. This motivation is clearly a long-run defense, since the strong law of large numbers only says that the average value will approach the expected value *in an infinitely long sequence of repeated experiments*.

Other examples of long-run defenses of rationality norms abound. For instance, the ‘Why ain’cha rich?’ argument is firstly proposed to defend evidential decision theory, which says that the one-boxer in the Newcomb’s Problem is richer than the two-boxer in the long run (Ahmed, 2014, p. 194; Hare & Hedden, 2016); new versions of the ‘Why ain’cha rich?’ arguments have been developed to undermine evidential decision theory (Gallow, 2021; Wells, 2017). One early defense of the Principal Principle says that following the principle allows us to choose a winning strategy in certain gamble when the gamble is played infinitely often (Mellor, 1982, p. 59). In philosophy of statistics, long-run considerations are used to motivate the Neyman-Pearson hypothesis test: as we use this inference rule infinitely often, the frequency of error will be low (Neyman & Pearson, 1933, p. 291). And the famous Bayesian Information Criterion in curve-fitting, which implies that we should select the simplest model that fits data, is motivated by the long-run consideration that the criterion will

select the model that maximizes posterior probability as data size tends to infinity (Schwarz, 1978).

Now, I don't mean to claim that long-run defenses are the *only* kind of defenses of the above norms of rationality, but only to emphasize that long-run considerations can reveal a lot of important features inherent in a method. Besides, it's true that the above examples don't involve exactly the same kind of long-run thinking—some focus on infinite repetition of *the same* evidence while others focus on infinite expansion of *growing* evidence; some focus on long-run *average* wealth while others focus on long-run wealth *creation*. However, they all suggest that questions like 'what if I do this every time' or 'what if I use this method infinitely often' could be relevant to our current, individual decision. Moreover, they seem relevant even if we know that we won't use the method infinitely often since we will die in finite time. Here is a helpful analogy. Recall the Dutch-book argument for probabilism, which says that being coherent keeps you from being dutch-booked. We may dislike this argument for many reasons, but it seems wrong to dismiss the argument simply by noting that there are no clever bookies around us. Or consider the money-pump argument for the transitivity of preference, which says that I could serve as a money pump if my preference is not transitive. Again, it seems wrong for me to dismiss this argument simply by noting that nobody around me is out there to exploit my incoherent preference. For the same reason, it seems wrong to dismiss all long-run defenses simply by noting that we won't live in the long run. The general lesson is that we may not want our beliefs or preferences to have certain features that are instantiated in certain settings, even if we know that those settings are not real.

### 5.3 What About the Short-Term Benefits of Rationality?

Here comes our third worry about the long-run: Perhaps it's true that we shouldn't dismiss convergence-theoretic epistemology for its focus on the long run, but the problem remains that convergence-considerations, as a kind of long-run considerations, seem too thin to do enough explanatory work that we want a theory of rationality to do. More specifically, we expect a theory of rationality to vindicate not only the long-term benefit of rationality but also the *short-term benefit*. Especially, we expect a theory to imply that rationality is beneficial even in a *fixed* rather than *growing* evidential situation. And convergence-theoretic epistemology can't accomplish the latter task.

In response to this worry, first recall that convergence-theoretic epistemology is highly successful in recovering norms of rationality like enumerative induction and simplicity; some initial attempts have been made to recover the reflection principle. And considering that there are still many unexplored modes of convergence,<sup>16</sup> convergence-theoretic epistemology can hope to recover further norms of rationality. If obeying those

---

<sup>16</sup> For instance, instead of requiring convergence to certainty in H if H is true and minimal confidence in H if H is false, we might require convergence to certainty in H if H is true and anything but convergence to certainty if H is false, or we may require convergence to minimal confidence if H is false and anything but convergence to minimal confidence if H is true. Moreover, when we know that the evidence stream is generated by some objective probability distribution (although we might not know what that distribution is), we may require that the convergence take place with high objective probability no matter what the underlying distribution is. (This kind of convergence is studied in PAC-learning in statistical learning theory.)

norms of rationality has short-term benefits, then convergence-theoretic epistemology can vindicate the short-term benefits of rationality. (If obeying those norms doesn't have short-term benefits, then perhaps we should rethink whether rationality necessarily has short-term benefit.)

Moreover, even if it turns out that convergence-theoretic epistemologists are unable to recover all rationality norms, they need not despair, because they can regard their project as an important *supplement* to, not a replacement of, the current decision-theoretic, accuracy-centered epistemology, which has been successful in explaining the short-term benefit of rationality for a wide range of rationality norms (i.e., benefits of obeying rational norms in fixed evidential situations).<sup>17</sup> Although I have explained the difference between the convergence framework and the current decision-theoretic framework, I don't mean to suggest that the two must be incompatible. For we might want *both* long-term benefits *and* short-term benefits in choosing a learning method, just like a company that aims at profit-maximization might consider both short-term profits and long-term profits in choosing a policy. Moreover, since rationality norms come in a variety of different forms, we shouldn't be surprised to find that they are truth-conducive for a variety of different reasons in the sense that some concern short-term benefits and others concern long-term benefits.

If we admit that both the long-term and short-term effects of a learning method matter for rationality, where does this leave of permissivism, dialectically speaking? My speculation is that it would bolster the case for permissivism, because it's likely to generate further sources of permissivity of rationality. For now we have additional trade-offs to make, if we assume that there are learning problems in which no method can maximize both long-term and short-term benefits (an assumption that I wish to confirm in future work). Then trade-offs will arise not only among long-term benefits themselves (since there are different desirable convergence modes, as I've mentioned above), but also *between* long-term benefits *and* short-term benefits. Some methods might achieve a high level of accuracy relative to the current evidential situation, but fail to improve accuracy as evidence grows once it reaches that level. Other methods might have to take much more evidence to achieve the same level of accuracy, but they are able to keep improving accuracy as evidence grows. An accuracy-maximizer might have to make a choice between the two kinds of methods, just like a company might have to choose between an unsustainable policy that brings huge short-term gain and a more sustainable policy that brings moderate and yet steady gain.

## 6. Conclusion

In this paper, I have proposed a new response to the value problem for permissivism: Rationality is valuable in pursuing accuracy because rational credence performs better with respect to convergence to truth, and this view is compatible with and even favors permissivism. I have argued that this convergence-theoretic epistemology has many interesting and attractive features. I hope my discussion has given permissivists a new way of explaining how rationality could be accuracy-conducive if it's permissive. In addition, I hope

---

<sup>17</sup> See Joyce (1998) and Pettigrew (2016). Note that all these attempts of recovering rationality norms don't assume the Expectation Constraint. So, it's unproblematic for permissivists to accept all of them.

that this paper will initiate further explorations of the convergence-theoretic epistemology, about what it can do and what it implies for other debates in epistemology.<sup>18</sup>

---

<sup>18</sup> This paper benefited greatly from discussion with David Thorstad, Selim Berker, Zach Barnett, Jie Gao, Jin Zeng, audiences at the First Asian Epistemologists Workshop, audiences at Brown University Epistemology Reading Group, and (especially) an anonymous referee for *PPR*.

## References

- Ahmed, A. (2014). *Evidence, Decision and Causality*. Cambridge: Cambridge University Press.
- Briggs, R. A. (2019), “Normative Theories of Rational Choice: Expected Utility”, *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2019/entries/rationality-normative-utility/>
- Dogramaci, S., & Horowitz, S. (2016). An argument for uniqueness about evidential support. *Philosophical Issues*, 26, 130–147. doi: 10.1111/phis.12078
- Earman J. (1992). *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge: MIT Press.
- Gaifman, H., & Snir, M. (1982). Probabilities over rich languages, testing and randomness. *Journal of Symbolic Logic*, 47, 495–548. doi: 10.2307/2273587
- Gallow, J. D. (2021). Riches and rationality. *Australasian Journal of Philosophy*, 99, 114–129. doi: 10.1080/00048402.2020.1716382
- Greaves, H., & Wallace, D. (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind*, 115, 607–632. doi: 10.1093/mind/fzl607
- Greco, D., & Hedden, B. (2016). Uniqueness and metaepistemology. *Journal of Philosophy*, 113, 365–95. doi: 10.5840/jphil2016113825
- Hare, C., & Hedden, B. (2016). Self-reinforcing and self-frustrating decisions. *Nous*, 50, 604–28. doi: 10.1111/nous.12094
- Hawthorne, J. (2018), “Inductive Logic”, *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2020/entries/logic-inductive/>.
- Horowitz, S. (2014). Immoderately rational. *Philosophical Studies*, 167, 41–56. doi: 10.1007/s11098-013-0231-6
- Horowitz, S. (2019). The Truth problem for permissivism. *Journal of Philosophy*, 116, 237–262. doi: 10.5840/jphil2019116516
- Joyce, J. M. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science*, 65, 575–603. doi: 10.1086/392661
- Juhl, C. (1997). Objectively reliable subjective probabilities. *Synthese*, 109, 293–309. doi: 10.1007/BF00413863
- Kelly, K. (1996). *The Logic of Reliable Inquiry*. Oxford: Oxford University Press.
- Kelly, K. 2007. How simplicity helps you find the truth without pointing at it. In M. Friend, N. Goethe, and V. Harizanov (Eds.), *Induction, Algorithmic Learning Theory, and Philosophy*. Dordrecht: Springer.
- Kelly, K. 2010. Simplicity, truth and probability. In P. Bandyopadhyay and M. Forster (Eds.) *Handbook of the Philosophy of Statistics*. Dordrecht: Elsevier.
- Kelly, K., Schulte, O., & Juhl, C. (1997). Learning theory and the philosophy of science. *Philosophy of Science*, 64, 245–267. doi: 10.1086/392550
- Kelly, T. (2014). Evidence can be permissive. In M. Steup, J. Turri, & E. Sosa (Eds.), *Contemporary debates in epistemology* (2nd ed., pp. 298–312). Chichester: Wiley-Blackwell.

- Kononenko, I., & Kukar, M. (2007). *Machine Learning and Data Mining*. Chichester: Horwood Publishing.
- Levinstein, B. A. (2017). Permissive rationality and sensitivity. *Philosophy and Phenomenological Research*, 94, 342–370. doi: 10.1111/phpr.12225
- Lin, H. (forthcoming). Modes of convergence: Steps toward a better epistemology of induction. *The Review of Symbolic Logic*.
- Meacham, C. J. G. (2019). Deference and uniqueness. *Philosophical Studies*, 176, 709–732. doi: 10.1007/s11098-018-1036-4
- Mellor, D. H. (1982). Chance and degrees of belief. In R. McLaughlin (ed.) *What? Where? When? Why?*, Dordrecht: Reidel.
- Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London, A*, 231, 289–337. doi: 10.1098/rsta.1933.0009
- Nielsen, M., & Wofsey, E. (2020). Speed-optimal induction and dynamic coherence. *British Journal of Philosophy of Science*, 1–19. doi: 10.1093/bjps/axz030
- Palmira, M. (forthcoming). Permissivism and the truth-connection. *Erkenntnis*, 1–16. doi: 10.1007/s10670-020-00373-7
- Pettigrew, R. (2013). A new epistemic utility argument for the principal principle. *Episteme*, 10, 19–35. doi: 10.1017/epi.2013.5
- Pettigrew, R. (2016). *Accuracy and the Laws of Credence*. Oxford: Oxford University Press.
- Plato. (1924). *Meno*. In *Laches, Pmtagoms, Meno, Euthydemus* (W. R. M. Lamb, Trans.) Cambridge: Harvard University Press.
- Putnam, H. (1963). Degree of confirmation and inductive logic. In P. A. Schilpp (Ed.) *The Philosophy of Rudolf Carnap* (pp. 761–83). La Salle: Open Court.
- Savage, L. J. (1972). *The Foundation of Statistics*. New York: Dover Publication, Inc.
- Schoenfield, M. (2014). Permission to believe: Why permissivism is true and what it tells us about irrelevant influences on belief. *Nous*, 48, 193–218. doi: 10.1111/nous.12006
- Schoenfield, M. (2015). Bridging rationality and accuracy. *Journal of Philosophy*, 112, 633–657. doi: 10.5840/jphil20151121242
- Schoenfield, M. (2019). Permissivism and the value of rationality: A challenge to the uniqueness thesis. *Philosophy and Phenomenological Research*, 99, 286–297. doi: 10.1111/phpr.12490
- Schulte, O. (1999a). The logic of reliable and efficient inquiry. *Journal of Philosophical Logic*, 28, 399–438. doi: 10.1023/A:1004443206028
- Schulte, O. (1999b). Means-ends epistemology. *British Journal for the Philosophy of Science*, 50, 1–31. doi: 10.1093/bjps/50.1.1
- Schulte, O. (2018). "Formal Learning Theory", *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/spr2018/entries/learning-formal/>>.
- Schultheis, G. (2018). Living on the edge: Against epistemic permissivism. *Mind*, 127, 863–879. doi: 10.1093/mind/fzw065
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–4. doi: 10.1214/aos/1176344136

- Thorstad, D. (2019). Permissive metaepistemology. *Mind*, 128, 907–926. doi: 10.1093/mind/fzy044
- Titelbaum, M. G. (2015). Continuing on. *Canadian Journal of Philosophy*, 45, 670–691. doi: 10.1080/00455091.2015.1124000
- Wedgwood, R. (2017). *The Value of Rationality*. Oxford: Oxford University Press.
- Weisberg, J. (2020). Could've thought otherwise. *Philosophers' Imprint*, 20.
- Wells, I. (2019). Equal opportunity and Newcomb's Problem. *Mind*, 128, 429–457. doi: 10.1093/mind/fzx018
- White, R. (2014). Evidence cannot be permissive. In M. Steup, J. Turri, & E. Sosa (Eds.), *Contemporary Debates In Epistemology* (2nd ed., pp. 312–323). Chichester: Wiley-Blackwell.
- Ye, R. (2019). The arbitrariness objection against permissivism. *Episteme*. 1–20. doi: 10.1017/epi.2019.42