



# A deference model of epistemic authority

Sofia Ellinor Bokros<sup>1</sup> 

Received: 8 May 2019 / Accepted: 24 August 2020  
© The Author(s) 2020

## Abstract

How should we adjust our beliefs in light of the testimony of those who are in a better epistemic position than ourselves, such as experts and other epistemic superiors? In this paper, I develop and defend a deference model of epistemic authority. The paper attempts to resolve the debate between the preemption view and the total evidence view of epistemic authority by taking an accuracy-first approach to the issue of how we should respond to authoritative and expert testimony. I argue that when we look at the debate through the lens of accuracy, it becomes clear that matters are more complicated than either the preemption view or the total evidence view are able to account for. Consequently, a deference model, outlined within a credence-based framework, does a better job of capturing the relevant phenomena, and explaining how we should update our beliefs in response to epistemically authoritative testimony.

**Keywords** Epistemic authority · Deference · Preemption · Expertise · Testimony · Accuracy first

## 1 Introduction

How should we respond to the testimony of experts, advisors, and more generally those who we consider to be in a better epistemic position than ourselves? In the recent philosophical literature on epistemic authority and expert testimony, we find two competing answers to this question.

According to *the preemption view*, we should completely defer to the testimony of those who are more knowledgeable than ourselves. Zagzebski (2012, 2013, 2016) has defended an influential account on which recognising somebody as an *epistemic authority* provides one with a preemptive reason to adopt their belief as one's own. In the words of Zagzebski, “[t]he fact that the authority has a belief *p* is a reason for me to believe *p* that replaces my other reasons relevant to believing *p* and is not simply

---

✉ Sofia Ellinor Bokros  
s.e.bokros@uni-koeln.de

<sup>1</sup> CONCEPT, University of Cologne, Cologne, Germany

added to them” (2013, p. 298). That is, instead of believing that  $p$  on the basis of one’s own reasons, when confronted with an epistemic authority, one is according to the preemptive account rationally required to adopt their belief on the basis of their epistemic authority. Preemptive views of epistemic authority have also been discussed and defended by Keren (2007, 2014a, b) and Constantin and Grundmann (2018).

The alternative view holds that, instead of constituting a *preemptive* reason, the authority’s testimony simply provides an *additional* reason upon which one can base one’s belief that  $p$ . On this view, one’s belief that  $p$  should be based on the total evidence that one possesses oneself, instead of being based solely on the belief of the authority. In other words, no matter how strong the authoritative reason might be, it should not replace or preempt the reasons that one already has for or against believing that  $p$ . As put by Jennifer Lackey, “the testimony of experts should always be regarded as a piece of evidence to be weighed with the other relevant evidence we have on the matter” (2018, p. 239). Lackey (2018), Dormandy (2018), Dougherty (2014) and Jäger (2016)<sup>1</sup> have all defended views in this spirit, although there are some important differences between their respective views. Following Constantin and Grundmann (2018), I will in this paper refer to this cluster of views as *total evidence views*, in reference to the core thesis that one’s beliefs ought to be based on one’s total evidence.

This paper attempts to shed light on the debate between the preemption view and the total evidence view by approaching the issue of epistemic authority from an accuracy-first perspective. I will argue that once we look at the debate through the lens of accuracy, we will see that matters are more complex than either the preemption view or the total evidence view fully account for.

In order to assess the debate from an accuracy-first point of view, I will transpose the debate on epistemic authority to the formal framework of accuracy-first epistemology, which makes use of credence functions to model doxastic states as well as an accuracy measure for measuring the accuracy of a credence function at a world. Within this framework, I will suggest that we conceptualise preemption as what in the context of formal epistemology has been labelled *complete epistemic deference* (Joyce 2007; Elga 2007). Complete epistemic deference takes place when an epistemic agent adopts the credence of her advisor with respect to  $p$  as her own, and thereby bases her own credence completely on the credal state of her advisor. In these terms, the issue at stake between the preemption view and the total evidence view would appear to be whether complete epistemic deference is the rational way of responding to epistemically authoritative testimony. In other words, the primary question I will try to address is whether complete epistemic deference results in the highest expected accuracy, or whether some other way of revising one’s credences in light of authoritative testimony would fare better in this regard.

To try and settle this matter, I will examine the highly controversial *track record argument*, which has been invoked in support of the preemption view. I will argue that when reformulated in terms of complete epistemic deference and expected accuracy, the track record argument does in fact provide us with strong initial support for a

<sup>1</sup> Although Jäger (2016) does not exclude that preemption is rational in some cases of epistemic authority, the positive account he defends (“socratic authority”) is one on which one ought not to preempt, but base one’s belief on one’s total reasons. The socratic approach can therefore be classified as a total evidence view.

deference model as the rational strategy. Secondly, I will address the problem of blind trust, which arguably appears to have been a decisive reason for some philosophers to reject the preemptive account in favour of a total evidence view, and which also appears to pose a problem for a deference model. I argue that contrary to what seems to be the consensus in the current literature, deference principles are not necessarily undermined by this objection.

Finally, in order to allow for uncertainties in our judgments of who is epistemically authoritative, I will introduce a weighted deference principle. Unlike the principle of complete epistemic deference, the weighted principle recommends partial deference in uncertain cases. However, this modification to the deference principle ultimately vindicates the total evidence view, while the preemption view is partially undermined. I thus conclude that insofar as accuracy is concerned, it appears that the total evidence view comes out on top in the vast majority of cases, while the preemption view only accounts for those cases in which one has absolute certainty of the other's authority.

The discussion will proceed as follows. In the next section, I will first outline the accuracy-first framework which I will adhere to and discuss some limitations of this framework. In section three, taking Zagzebski's account as a starting point, I provide a definition of epistemic authority in terms of expected accuracy that is suitable for our purposes. Section four presents the principle of complete epistemic deference as a way to formally think about preemption. Section five outlines the track record argument which provides a powerful initial justification for the principle of complete epistemic deference, while section six examines some exceptions to the track record argument. In section seven, I address the problem of blind trust. Section eight introduces a weighted deference principle, which allows us to account for potential uncertainties about our interlocutor's relative epistemic status. Lastly, in section nine, I briefly discuss the problem of competing authorities before I offer some concluding remarks in section ten.

## 2 Accuracy-first epistemology

In order to offer an additional perspective on the epistemic authority debate between the preemption and the total evidence view, this paper will make use of an accuracy-first approach to epistemology. Accordingly, before we get to the discussion of epistemic authority, there are a number of assumptions relating to the general framework of accuracy-first epistemology which should be made explicit.

Firstly, as suggested by its name, accuracy-first epistemology takes accuracy to be the sole epistemic good, and evaluates epistemic norms according to their ability to promote the good of having accurate doxastic states.<sup>2</sup> In order to do so, accuracy-first epistemology typically makes use of credence functions or "credences" to model doxastic states. The credence function of an epistemic subject *S* assigns a real number between 0 and 1 to any proposition *p* which *S* has a doxastic attitude towards. The number represents the credence which *S* has in *p* being true; a credence of 0 represents

<sup>2</sup> Accuracy-first epistemology also has a close affinity with *veritism*, which is the thesis that the sole or primary epistemic good is truth. For a generalised veritist approach to social epistemology, see Goldman (2001).

absolute certainty that  $p$  is false (there is no possibility that  $p$  could turn out to be true) and a credence of 1 represents absolute certainty that  $p$  is true (there is no possibility that  $p$  could turn out to be false). A credence of 0.5 in  $p$  represents a suspension of judgment with regards to  $p$ , that is,  $S$  thinks it equally probable that  $p$  could turn out to be true as that it could turn out to be false.

Accuracy-first epistemology can also be characterised as a version of epistemic utility theory. Epistemic utility theory applies tools from decision theory to epistemology in order to construct utility-based justifications for epistemic norms (see e.g. Pettigrew 2016, 2013a, b). However, instead of measuring the utility of actions, epistemic utility theory measures the epistemic utility of doxastic states. In accuracy-first epistemology, the epistemic utility is taken to be accuracy. It is thus assumed that a rational agent always attempts to adjust her credences so as to maximise her expected accuracy. More specifically, we can regard epistemic agents as if they were placing bets on the truth-value of each proposition to the value of their credence; the rational agent attempts to minimise his “losses” insofar as possible, by adopting the credence with the lowest expected inaccuracy (hence, being accurate is for the purposes of this paper simply a matter of not being inaccurate), as calculated by an accuracy measure (see Joyce 1998; Lewis 1980). An accuracy measure  $A$  is a function which takes a credence function  $c$  at a world  $w$ , and yields as its output a number  $A(c, w)$  which is a measure of the accuracy of the credence function at that world. There are many different measures of accuracy that can be used; in this paper I will be assuming the most standardly used measure, the Brier score, but as far as I am aware, assuming a different accuracy measure would not necessarily impact the argument. Relatedly, I will also assume that probabilism is true [which arguably follows from the accuracy-first framework (Joyce 1998)]. A further implication of this framework is that the model of epistemic authority discussed in this paper is only applicable to factual domains which are constituted by propositions with determinate truth values, i.e. which can be evaluated as either true or false, as otherwise it is not possible to appraise the accuracy of the agents’ doxastic states.

In addition, to keep things simple, I will assume that credences are sharp and that *the uniqueness thesis* is true (Elga 2010; Kopec & Titelbaum 2016). In other words, I will assume that for any proposition  $p$  and any set of evidence  $E$  there is only one maximally rational credence that a perfectly rational epistemic agent could have, which is a sharp value and not an interval, and any deviation from the unique maximally rational credence constitutes a deviation from rationality.

As with any models or frameworks, this approach to epistemology has both its advantages and its limitations. A clear advantage of using credences over the simple categorical framework of belief, disbelief, and suspension is that it allows for a more precise discussion of belief on authority, as the categorical framework cannot adequately model partial deference and other types of belief revision short of complete deference or preemption. However, since most of the discussion of epistemic authority thus far has relied on the categorical framework of belief,<sup>3</sup> it is not completely straightforward to transpose the insights and arguments from this debate to a credence-based

<sup>3</sup> Constantin and Grundmann (2018) also use credences to model doxastic states in their discussion of epistemic authority. However, they do not take an accuracy-first approach to the issue, which makes their project significantly different from mine.

model, as it is to some extent underdetermined what the preemption view and the total evidence view would be committed to once we move to this framework. There are many well-known problems with trying to reduce belief to credence (and vice versa) which have been extensively discussed in the literature, but in the interest of space, I cannot rehash that discussion here.<sup>4</sup> These difficulties naturally also apply to doxastic norms, and as far as I am concerned, there is no simple, principled way of moving from norms for categorical beliefs to norms for credences. Consequently, such a translation between the models has to proceed on a case-by-case basis, and one must be mindful of the fact that conclusions reached within one framework may not straightforwardly apply within the other. Nonetheless, it still seems to me a worthwhile project to discuss the issues surrounding epistemic authority within the context of accuracy-first epistemology, as it gives us an additional perspective on these issues. In addition, it allows us to shed light on some distinctions which are obscured by the categorical framework of belief, and bring out some further complexities of the phenomenon.

A further limitation which should be explicitly highlighted is the fact that, because accuracy-first epistemology takes accuracy to be the sole epistemic good, this comes at a cost of disregarding other epistemic goods and virtues. Admittedly, many of the participants in the debate on epistemic authority thus far have put great emphasis on epistemic and doxastic virtues other than accuracy. For instance, Zagzebski (2012, 2013) gives a central role to the virtue of epistemic conscientiousness in her discussion, which she defines as the “the quality of using our faculties to the best of our ability in order to get the truth” (2012, p. 48). Jäger (2016) underscores the importance of understanding, and central to his rejection of the preemption view is his emphasis on precisely understanding as “one of our main rational aims” (p. 180). Finally, Croce (2017) discusses a range of other intellectual virtues of epistemic authorities, such as sensitivity to the epistemic inferior subject’s needs and resources, intellectual generosity, and maieutic ability.

Because only accuracy is taken into account, the accuracy-first framework can only really incorporate the value of virtues such as understanding, conscientiousness, and sensitivity insofar as they contribute to accuracy. Needless to say, the benefits of these virtues are not always reducible to improvements in accuracy—if they improve accuracy at all. Effectively, this means that the benefits of these virtues are largely going to be left out of the following discussion.

Moreover, the current debate has implicitly been conducted with the aim of answering the question of how one, *all things considered*, ought to respond to epistemic authority. In contrast, the question which this paper will address is whether, from an accuracy point of view, a preemption or total evidence model of authority fares better. In other words, I am starting from the assumption that as long as epistemic agents are accurate, it does not make a difference whether they are conscientious, sensitive, or have a deepened understanding (although naturally, these things often go hand in hand). At the end of the day, however, there might be other virtues that trump accuracy, and this paper’s contribution to the debate is thus only one piece of the puzzle—albeit an important one.

---

<sup>4</sup> The literature on this topic is enormous, but see for instance Sturgeon (2008), Buchak (2014), Staffel (2017) and Jackson (2020) for discussion of some relevant issues.

In fact, I take it to be evident that many participants in the debate do take accuracy to be a highly important factor in the debate. What kicked off much of the debate was implicitly an argument from accuracy—namely Zagzebski’s appeal to the track record argument, as will be discussed in detail in section five. Answering this sub-question about accuracy is thus a crucial step towards answering the more general question of how we should respond to epistemic authority all things considered.

### 3 What is epistemic authority?

The next step in our discussion is to get more precise about how we should conceive of epistemic authority within the context of accuracy-first epistemology. To this end, I am going to take Zagzebski’s account of epistemic authority as my starting point, as arguably her account has been the most influential in the current debate.

Zagzebski (2013) bases her account of epistemic authority on Joseph Raz’s (1988) general principles of authority. She thus defines an epistemic authority as someone possessing a “normative epistemic power which gives me a reason to take a belief pre-emptively on the grounds that the other person believes it” (p. 296). On the basis of Raz’s account, she in total identifies four conditions which she takes to be constitutive of epistemic authority; Content-Independence, The Preemption Thesis, The Dependency Thesis, and The Justification Thesis (Zagzebski 2013, pp. 297–99):

**Content-Independence:** Under the assumption that the subject has reason to accept the authority as legitimate, the authority’s belief that  $p$  gives the subject a reason to adopt the belief that  $p$  that is not dependent upon the content of  $p$ . (That is, had the authority believed  $q$ , the subject would have had a reason to believe  $q$  instead.)

**Pre-emption Thesis for Epistemic Authority:** The fact that the authority has a belief  $p$  is a reason for me to believe  $p$  that replaces my other reasons relevant to believing  $p$  and is not simply added to them.

**Dependency Thesis:** If the belief  $p$  of a putative epistemic authority is authoritative for me, it should be formed in a way that I would conscientiously believe is deserving of emulation.

**Justification Thesis:** The authority of another person’s belief for me is justified by my conscientious judgment that I am more likely to form a true belief and avoid a false belief if I believe what the authority believes than if I try to figure out what to believe myself.

According to Zagzebski, we are justified in taking another epistemic agent to be an epistemic authority on the basis of the latter two conditions. Thus, it seems that it is in virtue of the latter two conditions being fulfilled that, according to Zagzebski, we have reason to attribute to them the normative power to generate preemptive as well as content-independent epistemic reasons.

Since one of the main issues which I am interested in in this paper is whether the preemption thesis for epistemic authority is indeed correct, I will exclude the conditions of having the power to generate preemptive or content-independent reasons from my definition of epistemic authority, as their inclusion would seem to beg the question which I am trying to answer. Hence, on the basis of Zagzebski’s account, we may take

as our starting point the general definition that an epistemic authority  $A$ , for me, is somebody who, on a given epistemic matter, forms their beliefs in a manner which I deem worthy of emulation, and who is more likely to form true beliefs with respect to said matter than I am independently of following  $A$ 's authority.

A natural way of expressing this idea in accuracy-first terms would be to think of the epistemic authority as somebody with a higher *expected accuracy* than oneself. While an accuracy measure allows us to measure an agent's accuracy with respect to a set of propositions given the truth values of those propositions (i.e. we can only measure accuracy with respect to a world), expected accuracy is measured given a probability function  $Pr$  over the proposition(s) which the agent whose accuracy we are measuring has a doxastic attitude towards, where  $(C, w)$  is an accuracy measure of a credence function  $C$  at a world  $w$ :

$$\sum_w Pr(\{w\}) \cdot A(C, w)$$

Hence, expected accuracy can be measured without already knowing which world we are in. This captures the idea that an epistemic authority is not necessarily someone who without fail has a more accurate belief than ourselves, but somebody who we think, in some objective sense, is *more likely* than ourselves to get at the truth. Accordingly, the probability function  $Pr$  with respect to which the accuracy should be taken is here to be conceived of in objective terms, which means that  $Pr$  will generally be unknown to epistemic subjects—i.e.  $Pr$  does not coincide with the epistemic subject's own credence function or subjective probabilities. Thus, to say that an epistemic authority  $A$  has a higher expected accuracy than some epistemic subject  $S$ , is to say that if we determined the expected accuracy of  $A$ 's credence function with respect to the objective probability function  $Pr$ , as well as  $S$ 's expected accuracy with respect to  $Pr$ , then we would find  $A$ 's expected accuracy to be higher than  $S$ 's.<sup>5</sup> In keeping with Zagzebski's original account, given that the only epistemic utility is accuracy, an epistemic agent with a higher expected accuracy is also by definition, at least in a thin sense, going to be "worthy of emulation" within the accuracy-first framework.<sup>6</sup> I would therefore propose the following definition of epistemic authority:

**Epistemic Authority:**  $A$  is an epistemic authority for  $S$  with respect to  $p$  iff  $S$  judges  $A$  to have a higher expected accuracy with respect to  $p$  than  $S$  takes herself to have independently of following  $A$ 's authority.

In addition, we can think of epistemic authority as grounded in an epistemic *superiority* relation, which we can define as follows:

<sup>5</sup> If one is sceptical of whether there is, or could be, such a thing as an objective probability function defined over all propositions, I would suggest that we can think of  $Pr$  in the following way. Imagine that there is an epistemic agent in the actual world, who possesses all the evidence that all the other epistemic agents in the actual world possess collectively and who has a perfectly unbiased ur-prior. Call her the absolute epistemic superior. Let  $Pr$  be whatever the absolute epistemic superior's credence function is.

<sup>6</sup> Zagzebski (2012) makes use of a much richer notion of being worthy of emulation (or imitation), specifically connected to the admiration of an exemplar, than it is possible for me to incorporate into my accuracy-oriented discussion. Thanks to an anonymous reviewer for drawing my attention to this.



**Epistemic Superior:** A is epistemically superior to S with respect to  $p$  iff A has a higher expected accuracy with respect to  $p$  than does S.

In this way, epistemic superiority is objectively defined whereas authority is a subjective relation; it is dependent on whether S makes the judgment that A is epistemically superior to herself. In other words, A acquires the status of an epistemic authority for S when S attributes epistemic superiority to A—this is also in line with Zagzebski's original account. Note, however, that the attribution of authority status is on this definition not necessarily a rational or accurate attribution, but can also be made irrationally or inaccurately. Further, since on this definition authority is relativised to agents, one might worry that a wildly inaccurate epistemic agent could rationally end up being an authority for somebody who is even more inaccurate (Keren 2014b, p. 71; Constantin and Grundmann 2018, p. 6). However, this possibility naturally rules itself out within the accuracy-first framework, as no rational agent would ever judge themselves to be less accurate than an indifferent credence function, because if they did, they would simply assign their own credences indifferently, as that would constitute an improvement in accuracy for them (Pettigrew 2016, p. 159).<sup>7</sup> Therefore, no one could rationally treat somebody who they took to have a lower expected accuracy than the indifferent credence function as an authority. This therefore imposes a minimum threshold on the reliability of authorities.

A potential weakness of the above definition of epistemic authority is that it provides little guidance on how to identify epistemic authorities. Because the expected accuracy is to be determined against the objective probability function  $Pr$ , finding the real expected accuracy is not something that epistemic agents are actually in a position to do, as the objective probabilities are almost never known.<sup>8</sup> That means that, although authority is defined in terms of expected accuracy, expected accuracy can in fact only be estimated, and usually only relative to some other agent (i.e. estimating somebody's expected accuracy in absolute terms is rarely feasible, but it might be a little more feasible to estimate who out of two agents has the higher expected accuracy). I shall therefore supplant this definition with a brief discussion of some epistemic qualities which actively contribute to a higher expected accuracy, and in virtue of which one might rationally judge somebody to be epistemically superior to oneself in this sense.

There appear to be two principal epistemic qualities in virtue of which one may count as epistemically superior to someone else. The first is by being epistemically superior with respect to *evidence*, and the second is by being epistemically superior with respect to *evaluation*. This distinction is intended to line up with the distinction Ned Hall (2004) draws between a *database expert* and an *analyst expert*, according to which the former “earns her epistemic status simply because she possesses more information,” and the latter “earns her epistemic status because she is particularly good

<sup>7</sup> An indifferent credence function is one which assigns credences according to the principle of indifference. The principle of indifference states that in the absence of any evidence, one should assign equal credence to each mutually exclusive and exhaustive possibility. See Pettigrew (2016, Part III) for discussion and different ways of formalising the principle of indifference.

<sup>8</sup> If the objective probabilities were known to an epistemic agent, the rational agent would naturally, by Lewis' principal principle, simply defer to the objective probabilities (provided that they do not have any inadmissible information), and wouldn't have any epistemic authorities (unless they took the authority to have inadmissible information).



at evaluating the relevance of one proposition to another” (p. 100).<sup>9</sup> Thus, I suggest two definitions:

**Evidence Superiority:** In regards to a claim  $p$ , an epistemic agent  $A$  is epistemically superior to  $S$  with respect to evidence iff  $A$  bases her credal state on a greater body of first-order evidence bearing on whether  $p$ , than  $S$  does.

**Evaluation Superiority:** In regards to a claim  $p$ , an epistemic agent  $A$  is epistemically superior to  $S$  with respect to evaluation iff  $A$  is better than  $S$  at evaluating and analysing evidence relevant to determining whether  $p$ .

In the first definition, ‘first-order evidence’ refers to the evidence that  $A$  has in their possession which bears directly on whether  $p$ .<sup>10</sup> In the second definition, being a better evaluator with respect to  $p$  should be understood as somebody who makes more accurate inferences from the available first-order evidence or is a more reliable reasoner, and thereby more often reaches the right conclusion about whether  $p$ . In a case where  $A$  is superior to  $S$  with respect to evidence, it seems *prima facie* plausible that,  $A$  and  $S$  being equally good evaluators, and all else being equal,  $A$ ’s judgment whether  $p$  is expected to be more accurate than  $S$ ’s judgment whether  $p$  because of the larger body of evidence bearing on  $p$  available to  $A$ . Likewise, in a case where  $A$  is superior to  $S$  with respect to evaluation, it seems highly plausible that,  $A$  and  $S$  having access to the same first-order evidence, and all else being equal,  $A$ ’s judgment whether  $p$  is expected to be more accurate than  $S$ ’s. Thus, the idea is that there are positive correlations between having more evidence and expected accuracy, as well as between being a better evaluator and expected accuracy.<sup>11</sup>

In the recent literature, an epistemic authority is usually taken to be superior with respect to *both* evidence and evaluation. Most paradigm examples of epistemic authorities in various domains—doctors, lawyers, scientists, historians—are considered to be authorities both because of their extensive and detailed first-order knowledge of their respective fields, and because of their superior abilities when it comes to evaluating and drawing correct conclusions from the data within their domain of expertise.

Strictly speaking, however, we should note that being superior with respect to evidence and/or evaluation is neither necessary nor sufficient for being epistemically superior *simpliciter* as I defined it above. For example, we could imagine counterexamples such as epistemic “authorities” who are superior to ourselves with respect to both evidence and evaluation but who somehow still end up with a lower expected accuracy than an indifferent credence function, or oracles who possess an extremely high

<sup>9</sup> Elga (2007, p. 479) also makes a similar distinction between deference to someone’s *information* and deference to their *judgment*.

<sup>10</sup> This definition is intended to exclude higher-order evidence about how pieces of first-order evidence bear on other pieces of first-order evidence and general evaluation skills, to avoid the distinction between evidence- and evaluation superiority collapsing. If such a distinction between evidence- and evaluation superiority ultimately cannot be upheld, nothing hinges on this distinction. Then we can simply conceive of epistemic superiority in terms of having more *evidence* in the most inclusive sense of the word.

<sup>11</sup> The former could be thought of as following from Carnap’s *Principle of Total Evidence* (Carnap 1947, p. 138), which states that a judgment of a probability based on a more inclusive body of evidence is always to be preferred to a judgment based on a less inclusive body of evidence. Of course, occasionally, a smaller body of evidence might be conducive to a more accurate judgment than a larger body, but as a general rule of thumb, the greater the amount of relevant evidence one possesses, the more accurate one is expected to be.

expected accuracy without neither more evidence nor superior evaluation skills. The latter notions are thus intended merely as useful heuristics for determining expected accuracy, not as perfect guides to it.

A further epistemic quality which is often mentioned in the literature on epistemic authority is that of being an *expert*. Experts are usually taken to be those epistemic agents who, in their respective domains of expertise, possess a wealth of knowledge and skills to make accurate judgments and assessments with respect to novel issues in those domains (Goldman 2001). However, it seems to me that the phenomenon of epistemic authority is much wider than that of expertise, and I have therefore not included the requirement of being an expert into the definition of epistemic authority.<sup>12</sup> To be exact, I take expertise to be neither necessary nor sufficient for epistemic authority. It is not *sufficient*, as one can be an expert in a domain without being epistemically superior to other experts in that domain, and it would in this case be odd to be considered an epistemic authority for those other experts (Dougherty 2014, p. 48). Nor is it *necessary*, as I find it plausible that one could count as epistemically authoritative on a wide range of everyday matters without necessarily being deserving of the label ‘expert’. For example, I am clearly epistemically superior to you with respect to my own mental states, or what I had for breakfast this morning, in virtue of which it seems that I should be epistemically authoritative for you on these matters, although it would in these cases sound odd to refer to me as an ‘expert’. This point has also been made by Croce (2019, pp. 14–15), who notes that a grandmother could figure as an epistemic authority for her young grandson on the topic of how fish breathe, despite only having elementary knowledge of zoology, because her grandson has no knowledge of the topic at all. Thus for her grandson, she qualifies as an epistemic authority in this case, despite not being an expert in any sense.

Another difference between authority and expertise is that authority appears to be more fine-grained than expertise. While expertise is commonly thought of as being relative to a domain, the definition of epistemic authority offered above defines epistemic authority relative to claims. In part, this is to keep our discussion of authority as straightforward as possible, as generalising over vaguely individuated domains obscures certain cases in which an authority might generally be in a better epistemic position with respect to a domain *D* than *S*, although *S* is better positioned with respect to some specific claim *p* within *D*. In addition, defining authority with respect to claims allows us to apply the model of epistemic authority to a broader range of cases, including more mundane cases of testimony, for which it might be difficult to define a domain of authority. This also seems to be in keeping with Zagzebski’s original account, as she leaves open the possibility of individual beliefs of an authority counting as authoritative.

Worth emphasising however is that the epistemic relationship between laypersons and experts still remains a paradigmatic instance of epistemic authority, because experts are generally those epistemic agents who we believe to be epistemically superior to ourselves with respect to those claims that fall within their domains of expertise.

---

<sup>12</sup> Arnon Keren (2007) also treats epistemic authority as applicable to ordinary testimony, thereby allowing it to be a much broader phenomenon than expertise.

As such, experts are very frequently epistemic authorities to other epistemic subjects with respect to the claims in their domains of expertise.

#### 4 Preemption as complete epistemic deference

Now that we have an idea of how we can think of epistemic authority within an accuracy-first framework, it is time to ask how we might think about the preemption view and the total evidence view, respectively, in accuracy-first terms.

A natural way to conceptualise preemption is as what formal epistemologists have referred to as *complete epistemic deference* (Joyce 2007, p. 190, Elga 2007, p. 479). S completely defers to A's judgment on  $p$  when she responds to A's judgment by updating accordingly, where  $A(p)$  is A's credence in  $p$ ,  $S^*$  is S's updated credence function, and  $S^*(p)$  is the resultant credence in  $p$  S should adopt:

$$\text{Complete Epistemic Deference: } S^*(p) = S(p \mid A(p) = x) = x^{13}$$

Complete epistemic deference seems analogous to preemption in two important respects, namely, it reflects both *deference of attitude* and *deference of reasons*.<sup>14</sup> Firstly, complete epistemic deference, just like preemption, has as a result that the epistemic subject comes to have the same doxastic attitude as the authority. If the authority has a credence of 0.9 in  $p$ , then that is also the credence which the epistemic subject will adopt, just like she would come to believe  $p$  if the authority believed  $p$  on the preemption view. Call this deference of attitude. Secondly, complete epistemic deference also captures the way in which the authority's belief that  $p$  is supposed to preempt or replace one's previous reasons for or against believing that  $p$ . For instance, when I completely defer to A on  $p$ , it does not matter if my credence in  $p$  before receiving the authority's testimony was 0.1 or 0.8, because that has no impact on the credence which I come to adopt. As can be seen, the epistemic subject's prior credence  $S(p)$  does not figure in the formula of complete epistemic deference. Nonetheless, when one defers, just like when one preempts, one's own evidence relevant to  $p$  does not somehow disappear—it simply no longer plays an active part in determining one's credence in  $p$ . In this sense, the reasons I previously had for adopting a credence of 0.1 or 0.8 in  $p$  can be said to be “normatively screened off” by the authority's judgment when I defer on  $p$ . Call this *deference of reasons*.<sup>15</sup>

<sup>13</sup> The principle of complete epistemic deference can also be regarded as a version of Bas van Fraassen's *Reflection Principle* (Elga 2007; van Fraassen 1984), although instead of deferring to a future, wiser self, we are deferring to an epistemic superior.

<sup>14</sup> I am grateful to an anonymous reviewer for making this distinction explicit.

<sup>15</sup> The word ‘deference’ here has two different senses; the sense of *deferring to someone or something*, as well as the sense of *suspending or putting to one side*. In ‘deference of attitude’ both senses are clearly applicable; one puts one's own credence to one side and one adopts the credence of the authority. In ‘deference of reasons’ only the second sense is clearly applicable, as one puts one's own reasons to one side, and one's immediate reason for having credence  $x$  in  $p$  is now that the authority has credence  $x$  in  $p$ . But since the authority's credence  $x$  in  $p$  is in turn based on the authority's reasons for  $p$ , one's own credence  $x$  in  $p$  is indirectly based on the authority's reasons for  $p$ , and thus arguably, one is also indirectly *deferring to the authority's reasons for  $p$* .

By contrast, it is much less clear what kind of principle could be invoked to formally represent the total evidence view. The only clear recommendation the total evidence view makes is that one should base one's credence on one's total evidence, as opposed to only basing it on the authority's credal state. On this broad characterisation, the total evidence view is only explicitly committed to denying deference of reasons, while remaining silent on the issue of how to adjust one's attitude. Among the defenders of the total evidence view there is certain variability in this regard, for instance, Dougherty (2014) is clearly committed to denying deference of attitude, whereas other authors (cf. Dormandy 2018; Lackey 2018; Jäger 2016) are much less committal on this aspect. It seems clear enough, however, that the total evidence view does not take deference of attitude to be required, even if it is not strictly ruled out that it might sometimes be appropriate. For this reason, I will for the purposes of this paper take the total evidence view to leave the issue of deference of attitude open.

In leaving the issue of deference of attitude open, the total evidence view would seem to leave room for a much wider range of possible credences which one could come to adopt upon receiving the authority's testimony. For example, if my prior credence in  $p$  is 0.1 and I subsequently become aware that the authority has a credence of 0.9 in  $p$ , the total evidence view could potentially allow me to adopt a range of different credences depending on what my total evidence supports. Perhaps I listen to the authority and become almost equally as convinced, such that I end up with a credence of 0.85, or perhaps I merely moderate my previous opinion and end up with a credence of 0.5—both ways of adjusting my credence in  $p$  seem compatible with the total evidence view, so long as the credence I adopt accords with my total evidence. It is thus to a large extent underdetermined what the total evidence view would in effect recommend once we move to a credence-based framework, if it would have anything to recommend at all. The only thing we could say for certain is that  $S(p)$  should somehow play a role in determining the credence which the epistemic subject comes to adopt, but there is no obvious formal rule which could serve as the formal analogue of the total evidence view. It is also not to be ruled out that different authors which I have characterised as all defending total evidence views would make rather different recommendations in this regard.

In what follows I will therefore initially focus on the principle of complete epistemic deference, and the question of to what extent complete epistemic deference is a rational way of responding to epistemic authority from an accuracy point of view. As will be seen in section eight, the principle of complete epistemic deference ultimately turns out to be inadequate, and in need of some refinement. To anticipate, the refinement will vindicate the total evidence view, but before we get there, we ought to get clearer on what preemption or complete epistemic deference might teach us about belief on authority.

## 5 The track record argument

There is a powerful argument in the recent literature, in support of preemption and against the total evidence view, which has become known as the *track record argument*.

The argument originally derives from Raz (1988), while its application to the epistemic domain is due to Zagzebski (2013). Here is the argument in Raz's original words:

Suppose I can identify a range of cases in which I am wrong more than the putative authority. Suppose I decide because of this to tilt the balance in all those cases in favour of its solution. That is, in every case I will first make up my own mind independently of the 'authority's' verdict, and then, in those cases in which my judgment differs from its, I will add a certain weight to the solution favoured by it, on the ground that it, the authority, knows better than I. This procedure will reverse my independent judgment in a certain proportion of the cases. [...] How will I fare under this procedure? If, as we are assuming, there is no other relevant information available, then we can expect that in the cases in which I endorse the authority's judgment my rate of mistakes declines and equals that of the authority. In the cases in which even now I contradict the authority's judgment the rate of my mistakes remains unchanged, i.e., greater than that of the authority. This shows that only by allowing the authority's judgment to preempt mine altogether will I succeed in improving my performance and bringing it to the level of the authority.

(Raz 1988, p. 68)

As should be clear, the track record argument is in effect a kind of accuracy argument—it argues that only by preempting, can the epistemic subject ensure that she is not less accurate in her judgment than the authority. Formulated explicitly in terms of expected accuracy and deference instead of preemption, we can spell out the track record argument as follows:

- (P1) When S completely defers to A on  $p$ , S's expected accuracy with respect to  $p$  will equal A's expected accuracy with respect to  $p$
- (P2) When S fails to completely defer to A on  $p$ , S's expected accuracy with respect to  $p$  will be lower than that of A
- (C) S will maximise her expected accuracy with respect to  $p$  if and only if S completely defers to A on  $p$

As (P1) appears to be indisputable, the only possibly controversial premise in this argument appears to be (P2). And indeed, whether the epistemically inferior subject could sometimes improve on the track record of the authority, i.e. whether (P2) should be endorsed, seems to be exactly what is at issue between those who endorse preemption and those who endorse a total evidence view. For instance, in response to Zagzebski, Keren (2014b, p. 74) argues that “[b]y suspending judgment in at least some proportion of the cases in which [the authority] believes  $p$  and I independently believe not- $p$  (or vice versa) [...] I can lower the probability of error not only below that of my own independent judgment, but also below that of the expert.” Hence, according to Keren, the layperson could sometimes do better than the authority by not deferring, and instead suspending judgment, which if true would refute (P2). In a similar spirit, Lackey (2018, p. 238) has also suggested that one could improve upon the accuracy of an authority by following a policy of partial deference. The idea would be to always defer, except when one is certain that the authority is wrong, in which case one should rely on one's own judgment. By following such a policy, it seems that

the result should be that the epistemic subject improves upon the performance of the authority, which is also in direct conflict with (P2). Therefore, it seems that whether the track record argument is sound hinges more or less entirely on the truth of (P2).<sup>16</sup>

Nevertheless, it seems to me that (P2) cannot cogently be denied in the manner in which Keren and Lackey argue. Although their suggestions have some intuitive plausibility, their alternative strategies both presuppose that the epistemic subject S has some reliable way of identifying cases in which the authority A is mistaken. That is, in order to improve upon the performance of A, S would need a method by which she could identify the cases in which A got it right and the cases in which A got it wrong, so that S could reverse the judgment of A only in those cases in which A got it wrong. If S simply reversed some of A's judgments arbitrarily, there would be no guarantee that the track record would be improved at all; rather, the contrary seems more likely. However, if S had such a reliable method for identifying cases in which A is wrong or too confident, it appears that S must in fact be in a *better* position epistemically than A with respect to those claims that S can reliably identify as mistaken. As such, this scenario appears to contradict the assumptions that we made about the superiority of A in the first place.

Under the assumptions that we have already made in section three about who counts as an epistemic authority, it simply doesn't seem possible for an epistemically inferior subject to be able to reliably assess when an authority is wrong. Surely, given that S has a lower expected accuracy than A, in cases where S and A disagree, it will be more likely that A got it right and that S got it wrong—not the other way around.

Moreover, if S really *did* have some reliable method of identifying claims made by A as correct or incorrect, we could not consider those claims to properly be part of A's domain of superiority, and thus not deferring with respect to those claims would be entirely compatible with the account, as A could no longer be considered to be an authority for S with respect to those claims. In fact, it seems that the roles then ought to be reversed; A ought to be consulting the judgment of S in order to improve the accuracy of their own judgment. Hence, S cannot coherently judge that A truly is epistemically superior, but that S herself is nonetheless in a good position to assess whether A is correct or incorrect.

Neither would Keren's suggested strategy of suspending judgment in those cases in which the epistemic subject and the authority disagree be better from an accuracy point of view. The simple reason for this is that suspending judgment would improve accuracy only if the expected accuracy of the authority was lower than the indifferent credence function (i.e. the credence function which, in a sense, suspends judgment on everything—at least if we think of suspending judgment as having a neutral credence which does not privilege any possibility over any other), and as stressed in section three, no rational epistemic subject should ever treat another epistemic agent with an

<sup>16</sup> It might seem that there is an additional manner in which S could outperform A, namely, by deferring to an alternative authority B. However, it is included in the definition of epistemic authority provided in section three that for A to be an authority for S, S does not have a higher expected accuracy independently of following A's authority. Thus, the possibility of S gaining a higher expected accuracy than A by deferring to an alternative authority B is already excluded. And consequently, if S can do better than A by deferring to B, A ceases to be authoritative for S when S defers to B. The issue of competing authorities will be more thoroughly addressed in section nine.

expected accuracy lower than the indifferent credence function as an authority. This means that, if S could reliably identify a range of cases over which A had an expected accuracy lower than the indifferent credence function (again, it is unclear how they would be in a position to do so), S ought not to treat A as an authority with respect to those cases.

As a final point of objection, one might worry that the upshot of the track record argument would be different if we spelled it out in terms of expected accuracy with respect to domains of authority instead of claims. If we were to define authority over domains instead of claims, this would open up for the possibility that although A is generally better positioned epistemically with respect to a domain D, S is better positioned with respect to some individual claims in D. In this case, S would potentially be able to reliably identify mistakes made by the authority with respect to those claims. This means that, if authority is defined over domains, we would also have to incorporate the additional assumption that the authority is better positioned with respect to every claim within D (or say that  $p$  falls outside of D), because otherwise the track record argument would not hold, and the objections made by Lackey and Keren would apply. But in either case, the final conclusion would still be that the track record argument does hold for the body of claims with respect to which the authority is better positioned. In sum, Keren's and Lackey's criticisms fail to apply to the track record argument as I have spelled it out here, as both criticisms rest on assumptions that are in conflict with our definition of epistemic authority.

## 6 Additional evidence

Although Keren's and Lackey's arguments against (P2) do not give us reason to reject the track record argument, there is a particular class of cases which seem to show that (P2) is not categorically correct. These are cases in which the authority A truly qualifies as an authority for S by having a higher expected accuracy with respect to the claim that  $p$  than S, but the inferior epistemic subject S would still be better off not simply deferring to A because of some additional piece of evidence that S possesses which the authority has not taken into account.

Consider the following scenario.

*Amateur Detective.* The police have been investigating a particularly difficult murder case for some time. To protect the ongoing investigation, the investigation records are not available to the public. The chief inspector announces to the public that she has a high credence that suspect A has committed the murder, although she cannot at the time disclose how she has come to that conclusion. An amateur detective who has been following the media reports closely has begun to look into the case, although he has no access to the materials gathered by the police and only limited information. In the process, the amateur stumbles on a piece of evidence which he can be certain has not been discovered by the police, which to the amateur seems to indicate that suspect B committed the murder.

Would it in this scenario be rational for the amateur detective to have a higher credence that suspect B committed the murder, or should he defer to the chief inspector, who



has a more complete picture of case, and adopt a higher credence that suspect A is guilty?

In line with what has been previously argued, a natural response would be that this depends on whether our subject, the amateur detective, has a higher expected accuracy than the putative authority, i.e. the chief inspector. If the amateur detective judges that because of the additional piece of evidence, the chief inspector does not have a higher expected accuracy on  $p$  than himself, then the chief inspector does not qualify as an authority for him on  $p$ , and he should not defer to the chief inspector. If the amateur detective instead makes the judgment that despite the additional piece of evidence, the chief inspector still has a higher expected accuracy than himself on  $p$ , then the chief inspector is an authority for him on  $p$ , and he should defer.<sup>17</sup>

However, Elga (2007) suggests a different model of updating for these kinds of scenarios, which conflicts with the deference model. Elga (2007, p. 480) proposes that in cases of this sort, a subject S ought to treat the authority A as a *guru* (where  $e'$  is S's additional piece of evidence):

$$\text{Guru: } S^*(p) = S(p \mid A(p|e') = x) = x$$

That is, S ought to believe what A would believe had A also taken S's additional evidence into account.<sup>18</sup> This means that the amateur detective should believe what the chief inspector would believe, had she also updated on the amateur detective's additional evidence. From the perspective of maximising expected accuracy, this proposal seems correct; surely, all else being equal, A's credal state with the additional evidence  $e'$  taken into account is going to have a higher expected accuracy than either A without  $e'$  or S with  $e'$ .

But if this line of reasoning is correct, we can now construct a counterexample to (P2) of the track record argument. If we stipulate that A without  $e'$  indeed has a higher expected accuracy than S with  $e'$ , and that S also makes the judgement that this is so, we now have a case in which S makes the judgment that A has a higher expected accuracy on  $p$  than S, and yet S should not completely defer to A on  $p$ , since S could do even better by following *guru*—by adopting the credence A would have had, had A also updated on  $e'$ . We thereby have a straightforward counterexample to (P2); the track record argument is not categorically correct.<sup>19</sup>

<sup>17</sup> Cases in which the subject is uncertain whether their interlocutor has a higher expected accuracy than themselves will be discussed in Sect. 8—for now we will just assume that either they make this judgement or they do not.

<sup>18</sup> However, if we are slightly arrogant and think that we are superior evaluators to our guru, perhaps one can turn the situation on its head and say that what we ought to believe is what we would ourselves believe if we had all of the guru's information. In fact, this seems to be relatively common real life epistemic behaviour; one gathers as much evidence as possible from the expert, and then makes one's own evaluation. However, this behaviour is only rational if one, by gathering the evidence from the expert, successfully transforms the situation such that oneself is now the epistemic superior. Needless to say, this is rarely going to be the case when laypersons interact with experts.

<sup>19</sup> It might be thought that the guru cases turn out to be cases in which A does not qualify as an authority for S, because by updating according to *guru*, S ends up with a higher expected accuracy than A, and so A is no authority for S. Largely, this is a definitional issue. In section three, I defined epistemic authority as follows: "A is an epistemic authority for S with respect to  $p$  iff S judges A to have a higher expected accuracy with respect to  $p$  than S takes herself to have independently of following A's authority." Much thus hangs on how we understand what it is to "follow" an authority; my intention was not for "follow" to be

What are the implications of this counterexample for the deference model? Although guru cases do qualify as genuine counterexamples to (P2) of the track record argument, I think they do rather little to threaten the general viability of the deference model. There are at least three reasons not to abandon the deference model just yet.

Firstly, for *guru* to outperform *complete epistemic deference*, it is absolutely crucial that  $e'$  is a piece of evidence that the authority has not already taken into account—which is not a very common occurrence. If it is not a new piece of evidence for the authority, and  $e'$  is already part of the authority's evidence base, then it seems that by following *guru*,  $e'$  would be double-counted (Kelly 2005; Keren 2007). Double-counting amounts to taking the same pieces of evidence into account twice over; once as part of the authority's evidence base, and once again as part of your own evidence, thereby illicitly updating on this evidence twice. The epistemic subject's resultant credence would thus not be proportional to the evidence available, as some evidence would be given double its weight. Arguably, this would make the subject's credences less accurate.

To clearly see why double-counting lowers expected accuracy, let's imagine for the sake of argument that a subject  $S$  possesses some evidence in favour of  $p$ , call this  $e^*$ , and on the basis of  $e^*$  takes  $p$  to be more probable than *not- $p$* .  $S$  then interacts with authority  $A$ , and finds out that  $A$  has a high credence in  $p$ , say 0.86, and that  $A$  has already taken  $e^*$  into account in her appraisal as a reason in favour of  $p$ . If instead of deferring to  $A$  on  $p$  and adopting  $A$ 's credence,  $S$  simply aggregates her own evidence for  $p$  with the judgment of the authority, then  $S$  could end up becoming more confident than the authority in  $p$ , adopting a credence higher than 0.86 in  $p$ . Now assume for reductio that this is a rational mode of updating for  $S$ . It would then have to follow, given uniqueness, that  $S$ 's higher credence has a higher expected accuracy than the authority's credence, because otherwise  $S$  should be better off simply deferring to the authority  $A$ . But if that were the case, it appears that  $A$  now is in an inferior epistemic position to  $S$ , and should in turn, in order to maximise her expected accuracy, update her credence so as to reflect that of  $S$ 's—but surely that would be unwarranted. From the perspective of the authority, learning that yet another layperson has come to the same conclusion as oneself, on the basis of evidence which one has already taken into account, does not provide one with a reason to become any more confident that  $p$  than one previously was.<sup>20,21</sup> The same argument can be made for cases in which  $S$  has

Footnote 19 continued

read as synonymous with complete epistemic deference. Plausibly, I think *guru* is still a way of “following” the authority.

<sup>20</sup> There are a couple of exceptions. Jäger's red cardinal case (2016, p. 175) and Constantin and Grundmann's mental math case (2018, 3, p. 19) are examples of cases in which an authority *could* rationally become more confident in their belief that  $p$  in virtue of finding out that a layperson has come to the same conclusion as themselves on the basis of evidence which they have already taken into account. That is because the layperson's agreement provides the authority with evidence that no perceptual error or performance error has occurred.

<sup>21</sup> Condorcet's jury theorem (Condorcet 1785) might be thought to tell us otherwise if we grant that epistemic agents are fallible in their evaluations. But this would presuppose 1) that we are adjudicating between a simple  $p$  or not- $p$  dilemma and 2) that the probability of the layperson coming to the correct verdict is higher than 0.5. If these conditions obtain, the expert has most likely already updated her belief in light of the majority opinion of the informed population. As such, one more layperson coming to the same verdict most likely has a negligible impact on the credal state of the expert.

some evidence against  $p$ , and ends up less confident than the authority in  $p$ . Hence, for *guru* to fare better than *complete epistemic deference*, it is absolutely crucial that  $S$ 's additional evidence is not already a subset of the authority's evidence. In all other situations the deference model still holds.<sup>22</sup>

Secondly, *guru* relies on an idealisation which makes it rather difficult to apply to real life scenarios. Namely, *guru* presupposes that the inferior epistemic subject is in a position to accurately estimate how the additional piece of evidence  $e'$  is going to affect the authority's credal state—but that is a condition which is rarely satisfied. Going back to our amateur detective, the amateur detective is unlikely to be in a position to judge whether the additional evidence he has discovered supports the theory that suspect B committed the murder. What *prima facie* seems to indicate that suspect B did it, might, given the chief investigator's overall evidence, turn out to better support the conclusion that suspect A did it, if it fits in with the chief investigator's best theory. In other words, the amateur detective might not be in a position to accurately judge whether the additional evidence makes it more likely, or less likely, that  $p$ . In such a scenario, it therefore seems that  $S$  could not improve on the judgment of  $A$  after all, as there is a serious risk of misevaluating the additional evidence and ending up with a less accurate prediction.  $S$  would therefore be better off simply deferring to  $A$  after all (indeed, the natural thing to do in such a scenario would be for the amateur detective to share  $e'$  with  $A$ , and then completely defer to  $A$  once  $A$  has updated on  $e'$  and made their revised belief known—but in the meantime, simply deferring to  $A$ 's judgment on  $p$  might still be the rational thing to do). Having additional evidence is therefore *on its own* not enough to justify not deferring; one additionally needs to be in a position to evaluate that additional evidence as competently as the authority would.

Finally, although *guru* conflicts with the deference model, it equally conflicts with the total evidence view. In a *guru* case, the subject is not required to reflect the credence of the authority, which means that such cases are counterexamples to deference of attitude, which is one aspect of the deference model. However, the central commitment of the total evidence view does not concern deference of attitude, but deference of reasons—namely, it is committed to opposing deference of reasons—and *guru* cases do not qualify as counterexamples to deference of reasons. According to *guru*, the subject should believe what  $A$  would believe given  $S$ 's additional evidence  $e'$ , but  $S$ 's prior evaluation of  $p$ ,  $S(p)$ , still plays no role in determining the credence which  $S$  should adopt. With the exception of  $e'$ , none of  $S$ 's own evidence or reasons for  $p$  should come into play, which makes *guru* still an instance of deference of reasons. *Guru* therefore conflicts with the total evidence view in this regard, and it seems fair to say that it is equally at odds with the total evidence view as with the deference model or preemption view.

In light of these considerations, I believe that despite the existence of these *guru*-type counterexamples to (P2), the deference model is still applicable to the vast majority of cases. The only instances in which the track record argument does not hold are cases in which the subject has some additional evidence which the authority has not already taken into account *and* is also in a position to evaluate that evidence as competently

<sup>22</sup> An alternative way of dealing with these counterexamples could perhaps be to make it a necessary condition on authority that for  $A$  to be an authority for  $S$  on  $p$ ,  $S$  must additionally make the judgment that  $S$ 's own evidence relevant to whether  $p$  is a subset of  $A$ 's evidence relevant to whether  $p$ .

as the authority would; cases which are ultimately going to be highly rare. Conceding that these cases are genuine exceptions to the deference model does not change the fact that the track record argument goes through in nearly all cases.

## 7 The problem of blind trust

Up until this point, it would appear that insofar as accuracy is concerned, the preemption view has the upper hand against the total evidence view. Nonetheless, there are still some additional considerations to be assessed, which reveal matters to be more complex than either the preemption view or the total evidence view account for.

The first issue which highlights some of these complexities is the problem of blind trust, to be addressed in the present section. A much discussed objection to the preemption view from the recent literature (Zagzebski 2013; Elga 2007; Lackey 2018; Jäger 2016), informally known as “the 4000-pills-case”, will help us draw these complexities into light. The case goes as follows.

*4000 pills.* A patient falls ill and pays a visit to their well-trusted doctor, who always in the past has given them great advice. After examining the patient’s symptoms and making a diagnosis as usual, the doctor gives the patient a treatment regimen: Every hour, take 4000 pills.

Undoubtedly, taking 4000 pills must be crazy advice! But if one adheres to either the preemption view or the principle of complete epistemic deference, it would seem that one is left with little choice but to follow the doctor’s recommendation; after all, the doctor knows best. Therefore, the deference requirement must be mistaken, as it appears to recommend blind trust in authorities, which is clearly irrational epistemic behaviour.

Is there a way for the proponent of the preemption view or the deference model to answer to this problem? According to Zagzebski (2013), one can reasonably disregard the doctor’s advice in this case, on the grounds that “your belief that you should not take so many pills is more likely to be true than your judgment that your physician is an authoritative guide to your health” (p. 302). Similarly, Elga (2007) argues that “[i]n realistic cases, one reasonably discounts opinions that fall outside an appropriate range. In addition, not even a perfect advisor deserves absolute trust, since one should be less than certain of one’s own ability to identify good advisors” (p. 483). In essence, I think these responses initially offered by Zagzebski (2013) and Elga (2007) are correct, albeit underdeveloped, hence why subsequent authors may have found their responses unpersuasive or ad hoc (Lackey 2018; Jäger 2016; Constantin and Grundmann 2018).

To develop a more detailed solution to the problem, we should first make a distinction between the judgment that  $p$ , i.e. the doctor’s judgment that one ought to take 4000 pills, and the judgment regarding the expertise of the doctor with respect to  $p$ , which we might call the superiority attribution judgment,  $q_{A(p)}$ . Thus, the deference model says that when we make the judgement that an authority  $A$  is superior with respect to the claim  $p$ , we should defer to  $A$ ’s credal state with respect to  $p$ . The 4000-pills-case thus constitutes a potential counterexample to the deference model insofar as it is a scenario in which  $S$  makes the judgement that  $q_{A(p)}$ , but it would be irrational for  $S$

to accept A's testimony on  $p$ . Granting that it truly would be irrational for the subject in such a scenario to accept  $p$ ,<sup>23</sup> what appears to be at issue is whether the deference model might still allow for the option of revising  $q_{A(p)}$ , instead of unconditionally accepting  $p$ .

The deference model can indeed allow for such an option. To see this, we should note that there are in fact two ways in which the deference requirement could be construed; either as a narrow-scope or as a wide-scope requirement.<sup>24</sup> On the narrow-scope formulation, the rational requirement has narrow scope. That is, once the subject S has attributed authority to A on  $p$ , it follows that S is rationally required to defer to A on  $p$ :

If S judges that  $q_{A(p)}$ , then rationality requires that [S defers to A on  $p$ ].

By contrast, on the wide-scope formulation, the rational requirement takes wide scope over the principle, requiring that the subject be such that she conforms to a conditional:

Rationality requires that [if S judges that  $q_{A(p)}$ , then S defers to A on  $p$ ].

While there is only one way to satisfy a narrow-scope requirement, i.e. by satisfying the consequent (in our case by deferring to A on  $p$ ), there are two ways in which one can comply with a wide-scope requirement. Either, one can satisfy the consequent part of the requirement by deferring to A on  $p$ , or, one can falsify the antecedent part of the requirement, by in our case withdrawing or revising the judgment that  $q_{A(p)}$ . Which of these two options is appropriate is left undetermined by a wide-scope requirement on its own, and what one ought to do can vary between situations. In other words, if we opt for a wide-scope formulation of the deference requirement, the deference model can in some cases still allow for the option of revising the judgment that  $q_{A(p)}$ .

Accordingly, by construing the deference requirement as wide-scope, it appears that the objection from blind trust can be avoided. In fact, on the wide-scope formulation, *4000 pills* no longer constitutes a counterexample to the deference model; by revising the judgment that  $q_{A(p)}$ , the deference requirement simply ceases to apply. On the wide-scope version, the 4000-pills-scenario by itself does not force one to accept the authority's testimony or violate the principle, as instead of violating the principle, one could simply revise one's judgment that  $q_{A(p)}$ .

In contrast, the narrow-scope version is clearly vulnerable to the objection from blind trust, as well as a host of related counterexamples on which authority is attributed to an epistemic agent who ought not to have that status. On the narrow-scope formulation, once S makes the judgment that  $q_{A(p)}$ , it follows that S ought to defer to A on  $p$ , irrespective of the conditions under which the judgment that  $q_{A(p)}$  was made, or whether it was made rationally or irrationally. That means that the narrow-scope version is not just vulnerable to *4000 pills*, but is generally unable to deal with cases in which the initial judgment that  $q_{A(p)}$  was not made rationally. On the narrow-scope version, if S makes the judgment that  $q_{A(p)}$  on the basis of A's charismatic appearance or under the influence of drugs, it would still follow that S ought to defer to A on  $p$ .

<sup>23</sup> The proponent of a preemptive account could also dispute this claim and argue that it is not in fact irrational to accept  $p$ ; see Constantin and Grundmann (2018) for an argument along these lines.

<sup>24</sup> See e.g. Way (2011) for a relatively recent exposition of the distinction and discussion of some related issues.

Accordingly, for the deference model to be at all plausible, it appears that we have little choice but to formulate the deference requirement as wide-scope.

That the objection from blind trust has sometimes been taken to be decisive against the preemption view might be because it has been tacitly assumed that the preemption thesis amounts to a narrow-scope requirement. For instance, Lackey (2018, p. 235) raises the worry that on the preemption view, it is not clear how subjects could even be in a position to tell that the testimony of an authority is outrageous. Since it is part of the preemption view that when S attributes authority to A on  $p$ , all of S's own evidence with respect to  $p$  is preempted or "normatively screened off", it appears that S is left without any resources to recognise A's testimony on  $p$  as crazy or outrageous. In other words, once S has attributed superiority to A with respect to  $p$ , it appears "too late", as it were, to revise  $q_{A(p)}$ .

There are a couple of points to be made in response to this objection. First off, a possible reading of this objection—without attributing this conflation to Lackey's original objection—trades on a conflation of the normative and the descriptive. The preemption view is a normative thesis, not a descriptive one. When preempted, the subject's own evidence or evaluation does not disappear (Constantin and Grundmann 2018, p. 9); it is "normatively screened off", which means that the subject ought not to rely on her own evidence or evaluation. Accordingly, the claim that S's evidence and evaluation is preempted with respect to  $p$  is not the claim that S loses all her epistemic resources with respect to  $p$ , rather, the claim would be that she *ought not* to rely on these resources with respect to  $p$ . Naturally, if the subject ought not to rely on these resources, she is in a sense no better off, as she would then be irrational in relying on these resources. But she is still in possession of these resources.

Secondly, and more importantly, the claim that S ought not to rely on these resources assumes a narrow-scope interpretation of the preemption view. In effect, the claim that once S has attributed authority to A on  $p$  it immediately follows that S ought not to rely on her own resources with respect to  $p$ , is just a restatement of the narrow-scope version of the requirement. Indeed, it is not obvious whether the wide-scope formulation or the narrow-scope formulation would be most faithful to Zagzebski's original account of the preemption view (although the general considerations against the narrow-scope formulation noted above would seem to indicate that it would be rather uncharitable to identify the original preemption view with the narrow-scope version). But since our primary aim here is to examine the merits of following a principle of complete epistemic deference, that question is to some extent orthogonal to our concerns; even if Zagzebski's original preemption view is most accurately thought of as narrow-scope, we can simply improve upon the original account and construe the deference model as a wide-scope requirement. On the wide-scope formulation of preemption and/or the deference requirement, it is not necessarily the case that S should disregard her own evidence and evaluation with respect to  $p$ , because in some cases what she should do is to rely on her own assessment of  $p$  and instead retract  $q_{A(p)}$ .<sup>25</sup>

<sup>25</sup> Further, it might be objected that even if a deference model can avoid the problem of blind trust, Zagzebski's account cannot allow for the option of revising  $q_{A(p)}$ , because then her criterion of Content Independence would be violated. But the Content Independence requirement states that "Under the assumption that the subject has reason to accept the authority as legitimate, the subject has reason to believe what

A general worry one might have about construing the deference requirement as wide-scope is that it makes the deference requirement too flexible, because it is not clear on what grounds one could be justified in revising  $q_{A(p)}$ . If instead of deferring one always has the option of revising  $q_{A(p)}$ , or effectively “demoting” the authority from their authority status, it would appear that the epistemic subject could simply choose to not defer whenever she pleases, or whenever the authority disagrees with her own views. In other words, under the assumption that the subject’s initial judgment that  $q_{A(p)}$  was made rationally, it seems doubtful whether the testimony itself could serve as appropriate grounds for withdrawing that judgment. Consequently, it might seem somehow problematic or even ad hoc, and in tension with the spirit of the deference model, to allow for the option of revising  $q_{A(p)}$ .

To answer to this worry, we should note that the fact that we do not yet have an account of when it is appropriate to revise one’s judgment that  $q_{A(p)}$  does not necessarily mean that “anything goes.” Naturally our superiority attribution judgments are also going to be subject to epistemic norms and rational constraints, and should not be made arbitrarily. Hence, if revising one’s judgment that  $q_{A(p)}$  is in our case ad hoc or otherwise irrational, it is irrational because one isn’t in this case justified in revising the judgment that  $q_{A(p)}$ —not because following a principle of epistemic deference is irrational or because it is misguided to formulate deference requirements as wide-scope.

As I see it, the real challenge posed by *4000 pills* is in answering the question of when, and on what grounds, one may rationally make or revise the judgment that some other epistemic agent is epistemically superior to oneself. Therefore, to adequately address the problem of blind trust, we need a rigorous decision procedure which will help us separate the cases in which it is rational to revise  $q_{A(p)}$  from the cases in which it is not, as well as more general criteria for assessing the competence of other epistemic agents. Consequently, the issue of how to identify epistemic authorities cannot be treated merely as an add-on to the issue of how we ought to respond to epistemic authorities once we have identified them; it is a highly important issue in its own right, which deserves its own treatment. Neither the preemption view nor the total evidence view adequately address this issue, since they are primarily views about how we should respond to epistemic authorities once we have identified them, and not views about how we identify them in the first place. It should be clearly acknowledged, therefore, that regardless of whether we endorse a deference requirement or not, we would need to offer some account here to complement our view.<sup>26,27</sup>

For obvious reasons, a full fledged account of this nature falls outside the scope of this paper. Nonetheless, I will make a couple more points on this topic before moving on.

---

Footnote 25 continued

the authority believes that is not dependent upon the content of the belief” (2013, p. 297, my italics), which also allows for a wide-scope interpretation which may be avoided by revising the antecedent.

<sup>26</sup> There is already some significant work done on this issue, see for instance Goldman (1999; 2001).

<sup>27</sup> A closely related problem is that of lying or insincere authorities. In addition to assessing the competence of authorities, we also need to assess their sincerity. For the purposes of this paper, I have taken the liberty to abstract away from the issue of insincerity in my discussion and simply assume that everyone is always sincere.



Firstly, as a preliminary suggestion in response to the revision problem: there is a potential solution to be found in our prior probabilities which I find relatively plausible. In keeping with the responses to *4000 pills* by Elga and Zagzebski, one candidate principle says that it is rational to revise one's judgment that  $q_{A(p)}$  when one's prior probability of the particular claim made by the authority is lower than the prior probability of oneself misjudging the authority. In other words, when one's prior probability of  $p$  is lower than one's prior probability of not- $q_{A(p)}$ , and both of these credences were rationally formed by the epistemic subject. This rule gives the right verdict even if one is relatively certain that the expert is epistemically superior; for instance, in the 4000-pills-scenario one might have a prior credence of 0.99 that she is genuinely epistemically superior on any claim in her domain of expertise, but nonetheless, one still has a higher prior credence of 0.999999 that one ought not to take 4000 pills. By epistemic conservatism, it then seems that revising one's belief in  $q_{A(p)}$  instead of  $p$  would be the rational response in that situation.<sup>28</sup>

Secondly, the deference model is entirely compatible with the view that one ought to use one's *total evidence* when deciding whom to treat as an epistemic authority. Endorsing a deference principle does not preclude one from basing one's judgment that  $q_{A(p)}$  on the total evidence that one has available, or even on deference to someone else whom one trusts. In fact, it would seem highly suspicious if one were to base one's judgment that  $q_{A(p)}$  on deference to A—if anything, *that* would be an instance of blind trust! To avoid such circularity, it is thus inevitable that one's judgment that  $q_{A(p)}$  be based on one's own assessment, or deference to an independent authority.

Third, it does not follow from the deference requirement that any evidence in one's possession that is relevant to determining whether  $p$  is rendered useless with respect to  $q_{A(p)}$  (or any other claim for that matter). The manner in which one's own reasons are "normatively screened off" does not amount to those reasons becoming *generally* inactive or unusable (Constantin and Grundmann 2018, p. 9; Jäger 2016, p. 175). Rather, when deferring on  $p$ , one's reasons for  $p$  are only screened off with respect to  $p$  itself.

In some sense, it could seem that allowing deference to be conditional on other beliefs that are based on one's total evidence in this way shows that the total evidence view is correct. There is some truth to this observation; whether one comes to defer to the authority depends on whether one takes them to be epistemically superior, and whether one takes them to be epistemically superior is most likely going to be based on one's total evidence—which possibly includes how plausible one finds the claims that they make. It should therefore be conceded that *indirectly* one's credence in  $p$  is in some sense always based on one's total evidence, also on the deference model. Indeed, what the 4000-pills-case appears to show is that *who* to believe and *what* to believe cannot be decided entirely independently, but to some extent needs to be determined holistically. Yet, this insight does not invalidate the deference model. What the deference requirement says is that under certain conditions, with respect to certain claims, it is not rational to directly base one's credences on one's total evidence; it does not demand deference "all the way down", as it were.

<sup>28</sup> One could also think of one's credence in the possibility that one has misidentified the authority A as setting a variable minimum threshold which one's prior credence in  $p$  needs to meet for deference to A on  $p$  to be rational.

## 8 Uncertainty about the superiority of one's interlocutor

The preceding discussion points us to a further problem for the principle of complete epistemic deference, which will be the topic of the present section. So far in the discussion I have assumed that judgments about other epistemic agents' competences are categorical, all-or-nothing type judgments; either we take A to be epistemically superior, or we do not take A to be epistemically superior. It might therefore seem that the conclusions drawn thus far are in fact only valid for cases in which we have complete certainty or trust in our interlocutor's epistemic superiority; in other words, cases in which we assign credence 1 to  $q_{A(p)}$ .

But what if one's credence in  $q_{A(p)}$  does not amount to complete certainty? What if one has a credence in  $q_{A(p)}$  of 0.95, or 0.7? As already noted, we are not infallible when it comes to identifying our epistemic superiors, or even peers. As reasonable individuals, we are also aware of our own fallibility in this regard, and thus we should hardly ever take ourselves to be so certain as to have credence 1 in anyone's epistemic superiority. In addition, there is also a large class of cases which fall into the grey area between being cases of epistemic authority and cases of perfect peer disagreement, in which one might find it reasonably likely that one's interlocutor is epistemically superior with respect to  $p$ , although one is nowhere near certain of this judgment. How should one adjust one's credences in light of one's interlocutor's doxastic state in such uncertain situations?

Plausibly, how one should update in a situation in which one does not have credence 1 in one's interlocutor being epistemically superior mirrors how one should update in a situation in which one acquires some new evidence without assigning credence 1 to that new evidence (i.e. by Jeffrey Conditionalisation (see Jeffrey 1965, chapter 11)). In other words, it seems that one could take a weighted linear average<sup>29</sup> of the credence in  $p$  one would adopt were one completely certain of A's superiority and the credence in  $p$  one would think to be correct in the event that A is not epistemically superior, i.e. one's own prior credence in  $p$ . In effect, this would amount to *partial* epistemic deference, where one's degree of deference is proportional to how probable one judges it to be that A is epistemically superior<sup>30</sup> :

$$\text{Weighted Epistemic Deference: } S^*(p) = A(p) S^*(q_{A(p)}) + S(p) S^*(\neg q_{A(p)})$$

What can now be noted is that in those cases in which one's credence in the other being epistemically superior is less than 1, one's resultant credence is no longer a reflection of the credence of the epistemic superior. Consequently, one's resultant credence is now a product of one's independent opinion and one's interlocutor's opinion, each weighted according to how probable one finds it that the other is epistemically superior.

<sup>29</sup> Dougherty (2014, p. 56) also suggest that we should use a weighted linear average for weighing the opinions of authorities. However, the only consideration he cites in support of this type of weighing is that "authority clearly comes in degrees" (p. 57), whereas my recommendation here is based on considerations of accuracy. In addition, the weightings on Dougherty's approach are based on to what degree the authority is an expert; on my approach they are based on the probability the epistemic subject assigns to the proposition that the interlocutor is epistemically superior to them.

<sup>30</sup> We are here using  $S^*(q_{A(p)})$  and  $S^*(\neg q_{A(p)})$  instead of  $S(q_{A(p)})$  and  $S(\neg q_{A(p)})$  to avoid the objection from blind trust discussed in the previous section.

However, in the event that  $S^*(q_{A(p)}) = 1$ , we do still get full deference. For a rational agent conforming to probabilism, when  $S^*(q_{A(p)}) = 1$ , then  $S^*(\neg q_{A(p)}) = 0$ , and  $S(p)$  will have no impact on  $S^*(p)$ , just as before. We can thus use this formula to model all cases and get the same results; the principle of complete epistemic deference falls out of this more complex principle when  $S^*(q_{A(p)}) = 1$ .

An important caveat, however, is that just as in the previous sections, we are still assuming that the inferior epistemic agent does not have any additional evidence to the superior (even if it is uncertain who out of the two is the superior). If the inferior epistemic agent has some evidence that the superior does not, this type of updating by weighted linear averaging will be inadequate, just like complete epistemic deference fails in the guru cases discussed in section six (see also Bradley 2017).

If this model is correct, it would seem to undermine the preemption view, because this shows that unless one is completely certain of the other's authority, one's own prior opinion *is* going to play some role in determining one's posterior credence after all. That is,  $S(p)$  partly determines which credence  $S$  adopts upon receiving  $A$ 's testimony, which means that there is no longer a deference of reasons taking place. Neither is there a deference of attitude, as the credence adopted by  $S$  can differ from  $A$ 's. The total evidence view is thereby vindicated, as it is now truly accurate to describe the situation as one in which one does not defer one's reasons, but one's credence is based on the total evidence, including one's own evidence and opinions which one had prior to consulting the "authority".

In practice, however, it seems that the influence of one's prior opinion ought in most cases when one interacts with experts be so small that it would appear almost negligible. That is because once one's credence in the other's superiority is sufficiently high, one's previous opinions will receive such a small weighting that they barely make any difference to the final credence one should adopt. Arguably, most situations in which laypersons interact with experts are going to be situations in which such high credences in their superiority will be perfectly appropriate, and having credences that are lower would generally seem unwarrantedly arrogant or sceptical. The same would generally go for instances of ordinary testimony—if you tell me about what you had for breakfast this morning, it seems that my credence that you are epistemically superior to me on this matter ought to be extremely close to 1. The situations in which my credence in your superiority are in the mid-ranges, say ranging between 0.1 and 0.9, are plausibly more typically thought of as types of *peer disagreements* loosely defined, or situations in which we are at least roughly peers<sup>31</sup> (a *perfect* peer disagreement would be a situation in which my credence in your superiority is 0.5<sup>32</sup>). In practice, a typical expert-layperson interaction will include, or even require, a nearly full deference of attitude to the expert's opinion.

That being said, although within the accuracy-first framework we would hardly ever have epistemic reason to completely defer to an authority in the sense of giving their

<sup>31</sup> This principle thus also tells us how we ought to respond to those who we in all likelihood judge to be our epistemic inferiors—i.e., if I have a credence of 0.9 in you being epistemically inferior to me, I should, if I'm rational, also have a credence of 0.1 that you are epistemically superior, which means that I should give your view a weighting of 0.1.

<sup>32</sup> Note that the principle of weighted epistemic deference also implies an *equal weight view* for perfect peer disagreements in which the peers possess the same evidence.

credence in  $p$  the full weight (as we should plausibly never be entirely certain of our interlocutor's superiority), it is not to be ruled out that we sometimes have other kinds of reasons—perhaps moral reasons—to treat the testimony of our interlocutors in this way. A question for further exploration is thus if we ever ought to behave doxastically *as if* we had credence 1 in the other's epistemic superiority; that is, if it is sometimes appropriate to give one another absolute epistemic authority, regardless of whether we in fact have credence 1 in one another's epistemic superiority.

## 9 Competing authorities

Before I conclude the discussion, I wish to add some remarks on one final problem; namely, how one should respond epistemically when confronted with competing authorities. There is already a large body of literature on the problem of expert disagreement, and I will not be able to do justice to that literature here. Accordingly, what I wish to briefly do in this section is to draw out some of the implications of the deference model for how one ought to respond to competing authorities, and also show that a deference model is not necessarily going to be undermined by the existence of multiple authorities.

Firstly, the existence of more than one authority might seem to pose a problem for a deference model, as it would seem impossible for an epistemic agent to defer fully to two or more authorities at once, unless the authorities have the exact same credence (Jäger 2016; Bradley 2017; Gallow 2018). Yet, it seems to be a perfectly coherent possibility that for an epistemic subject  $S$  there could be more than one epistemic authority, i.e. there could be multiple epistemic superiors who  $S$  judges with complete certainty to have a higher expected accuracy than herself, but who have wildly different credences from each other.

The solution, which falls out of our definition of epistemic authority from section three, is that one ought to defer to the judgment of the authority which one takes to be the most superior. Whoever is judged to have the highest expected accuracy is the authority one ought to defer to (again assuming that one doesn't take oneself or any of the other authorities to have any additional evidence which  $A$  has not taken into account<sup>33</sup>). The reason why one can in this case completely defer to the authority with the highest expected accuracy, and not worry about the others, is that from the perspective of the agent  $S$ , once  $S$  defers to the most superior authority, let's call him authority  $A$ ,  $S$  will no longer judge that either authority  $B$  or authority  $C$  have a credal state with a higher expected accuracy than her own, and thus  $B$  and  $C$  will cease to be authoritative for  $S$ . On the other hand, if  $S$  initially only encounters and defers to  $B$ ,  $S$  will still defer to  $A$  once  $S$  is confronted with  $A$ 's epistemic superiority. Another way of putting this point is that if  $S$  judges  $A$  to be superior to  $B$  and  $C$ ,  $S$  also thinks that  $B$  and  $C$  ought to defer to  $A$ . And because  $B$  and  $C$  really ought to defer to  $A$ ,  $S$  can disregard  $B$  and  $C$ 's judgments and simply defer to  $A$ . So, in effect, in order to

<sup>33</sup> If we drop this assumption, we would be ignoring additional evidence by simply deferring to the most superior authority. This situation would thus call for some mode of updating analogous to the Guru cases above; strictly speaking, we ought to believe what the most superior authority would believe had they also considered the additional evidence.

maximise accuracy, each epistemic agent ought to defer to whoever has the highest expected accuracy. If S judges that A is the one who has the highest expected accuracy, S defers to A, and consequently, S also believes that everyone else ought to defer to A in order to maximise their accuracy.<sup>34</sup>

What about cases in which it is not transparent who the best expert actually is, and one believes that A, B, and C are equally likely to have the highest expected accuracy with respect to  $p$ ? Assuming that A, B, and C possess the same evidence, it follows from the principle of weighted deference which I endorsed above that one should give each expert the weight which corresponds to how probable one judges it to be that they are the most superior. In this case, then, one should give A, B, and C equal weight, i.e. 0.333.... One will ultimately end up with the exact same credence regardless of whether one encounters them all at once, or if one defers to them in the order in which one meets them, because the assessment of their relative superiority to one's own credal state, and thus the weights one assigns, will change accordingly as one defers and improves the expected accuracy of one's own credal state. For example, if S first encounters authority A, S will at that point defer fully to A, because A is absolutely certain to be epistemically superior to S and has no competition. Then, upon encountering B, S will judge A and B to be equally likely to be the most superior, and thus have a credence of 0.5 in  $q_{B(p)}$  and update accordingly. Then, once S encounters C, S will have a credence of 0.333... in  $q_{C(p)}$ , since the likelihood of her own credal state having the highest expected accuracy now equals the likelihood of A and/or B's credal state being the most superior out of A, B and C. After updating on the opinions of A, B, and C, S will by her own lights have a credal state with a higher expected accuracy than either A, B, or C with respect to  $p$  and so her credence stabilises. Note that the same updating model works even if one does not give the different experts equal weight, provided that the epistemic agent consistently updates her belief about the relative superiority of her own credal state by conditionalising on the relative superiority of the authorities to whom she has already deferred.

## 10 Conclusion

In this paper I have examined the phenomenon of epistemic authority from an accuracy-first perspective. By so doing, I have attempted to shed some light on the debate between the preemption view and the total evidence view. While an updated version of the track record argument formulated in terms of deference and expected accuracy provides strong initial support for a preemptive view, once additional factors and uncertainties are taken into account, a total evidence view seems closer to the mark. A general account of how we should respond to epistemic authorities can instead be supplied in terms of a principle of weighted epistemic deference, which also has implications for how we ought to respond to our epistemic peers and inferiors. Ultimately, the underlying weighted deference principle which I have defended vindicates the total evidence view in the great majority of cases, in which we are not completely certain about our interlocutor's epistemic status relative to our own, while the preemption

<sup>34</sup> See Gallow (2018) for some formal results which are consistent with this solution.

view appears correct in cases in which we have full certainty of our interlocutor's epistemic superiority.

Nevertheless, an additional takeaway of the discussion is that deference principles by themselves offer no guidance on *when* one is justified in treating another epistemic agent as epistemically authoritative. Naturally, this points us to further questions; how can we reliably assess other epistemic agents and identify our epistemic superiors? On what normative grounds can we criticise individuals who defer to epistemic authorities who do not deserve their authority status, or who fail to defer to their true superiors? I have had very little to say about these issues here. Similarly, I have had nothing to say about epistemic desiderata other than accuracy, such as understanding or intellectual independence, which could potentially alter the picture of how one in practice should respond to epistemic authorities. While a deference model appears largely correct in the idealised cases discussed here, admittedly, real life is rarely as clear cut.

**Acknowledgements** Open Access funding provided by Projekt DEAL. Many thanks to Lisa Benossi, Sven Bernecker, Charlie Blunden, Daniel Cohnitz, Thomas Grundmann, Stefanie Grüne, Giulia Napolitano, Jakob Ohlhorst, Francesco Praolini, Luis Rosa, Paul Silva, Eyal Tal, the audiences at GAP.10 in Cologne and at the Graduate Conference in Theoretical Philosophy at the University of Amsterdam in 2018, as well as to three anonymous reviewers from Synthese, for helpful feedback on earlier versions of this paper.

**Funding** This research was supported by the Cologne Center for Contemporary Epistemology and the Kantian Tradition (CONCEPT), funded by the Alexander von Humboldt Foundation and the Federal Ministry of Education and Research.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Bradley, R. (2017). Learning from others: Conditioning versus averaging. *Theory and Decision*, 85(1), 5–20.
- Buchak, L. (2014). Belief, credence, and norms. *Philosophical Studies*, 169, 285–311.
- Carnap, R. (1947). On the application of inductive logic. *Philosophy and Phenomenological Research*, 8, 133–148.
- Constantin, J., & Grundmann, T. (2018). Epistemic authority: Preemption through source sensitive defeat. *Synthese*. <https://doi.org/10.1007/s11229-018-01923-x>.
- Croce, M. (2017). Expert-oriented abilities vs. novice-oriented abilities: An alternative account of epistemic authority. *Episteme*. <https://doi.org/10.1017/epi.2017.16>.
- Croce, M. (2019). On what it takes to be an expert. *The Philosophical Quarterly*, 69(274), 1–21.
- de Condorcet, M. (1785). *Essay on the Application of Analysis to the Probability of Majority Decisions*. Paris: Imprimerie Royale.
- Dormandy, K. (2018). Epistemic authority: Preemption or proper basing? *Erkenntnis*, 83(4), 773–791.
- Dougherty, T. (2014). Zagzebski, authority, and faith. *European Journal for Philosophy of Religion*, 6(4), 47–59.
- Elga, A. (2007). Reflection and disagreement. *Noûs*, 41(3), 478–502.
- Elga, A. (2010). Subjective probabilities should be sharp. *Philosophers' Imprint*, 10(5), 1–11.

- Gallow, J. D. (2018). No one can serve two epistemic masters. *Philosophical Studies*, 175(10), 2389–2398.
- Goldman, A. I. (1999). *Knowledge in a social world*. Oxford: Oxford University Press.
- Goldman, A. I. (2001). Experts: Which ones should you trust? *Philosophy and Phenomenological Research*, 63(1), 85–110.
- Hall, N. (2004). Two mistakes about credence and chance. *Australasian Journal of Philosophy*, 82(1), 93–111.
- Jackson, E. G. (2020). The relationship between belief and credence. *Philosophy Compass*. <https://doi.org/10.1111/phc3.12668>.
- Jäger, C. (2016). Epistemic authority, preemptive reasons and understanding. *Episteme*, 13(2), 167–185.
- Jeffrey, R. (1965). *The logic of decision*. Chicago: University of Chicago Press.
- Joyce, J. M. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science*, 65(4), 575–603.
- Joyce, J. M. (2007). Epistemic deference and the case of chance. In *Proceedings of the Aristotelian Society*, CVII (pp. 187–206).
- Kelly, T. (2005). The epistemic significance of disagreement. In T. Gendler & J. Hawthorne (Eds.), *Oxford Studies in Epistemology* (Vol. 1, pp. 167–196). Oxford: Oxford University Press.
- Keren, A. (2007). Epistemic authority, testimony, and the transmission of knowledge. *Episteme*, 4(3), 368–381.
- Keren, A. (2014a). Trust and belief: A preemptive reasons account. *Synthese*, 191(12), 2593–2615.
- Keren, A. (2014b). Zagzebski on authority and preemption in the domain of belief. *European Journal for Philosophy of Religion*, 4, 61–76.
- Kopec, M., & Titelbaum, M. G. (2016). The uniqueness Thesis. *Philosophy Compass*, 11(4), 189–200.
- Lackey, J. (2018). Experts and peer disagreement. In M. A. Benton, J. Hawthorne, & D. Rabinowitz (Eds.), *Knowledge, belief, and god: New insights in religious epistemology* (pp. 228–245). Oxford: Oxford University Press.
- Lewis, D. K. (1980). A subjectivist's guide to objective chance. In R. C. Jeffrey (Ed.), *Studies in inductive logic and probability* (pp. 83–132). California: University of California Press.
- Pettigrew, R. (2013a). Epistemic utility theory. *The Reasoner*, 7(1), 10–11.
- Pettigrew, R. (2013b). Epistemic utility and norms for credences. *Philosophy Compass*, 8(10), 897–908.
- Pettigrew, R. (2016). *Accuracy and the Laws of Credence*. Oxford: Oxford University Press.
- Raz, J. (1988). *The morality of freedom*. Oxford: Clarendon Press.
- Staffel, J. (2017). Accuracy for believers. *Episteme*, 14(1), 39–48.
- Sturgeon, S. (2008). Reason and the grain of belief. *Noûs*, 42(1), 139–165.
- van Fraassen, B. C. (1984). Belief and the will. *The Journal of Philosophy*, 81(5), 235–256.
- Way, J. (2011). The symmetry of rational requirements. *Philosophical Studies*, 155(2), 227–239.
- Zagzebski, L. (2012). *Epistemic authority: A theory of trust authority and autonomy in belief*. Oxford: Oxford University Press.
- Zagzebski, L. (2013). A defense of epistemic authority. *Research Philosophica*, 90(2), 293–306.
- Zagzebski, L. (2016). Replies to Christoph Jäger and Elizabeth Fricker. *Episteme*, 13(2), 187–194.