Preemptionism and its Reliabilistic Assumption: A Bayesian model[*]

Christian J. Feldbacher-Escamilla

Summer 2025

1 Introduction

[57] *Preemptionism* posits that, when an epistemic authority or expert contra a layperson holds a belief or disbelief in a proposition, the layperson should entirely supplant any reasons for or against that proposition held by the epistemic subject under the authority's purview. The result will be deference to the epistemic attitude of the authority. This account has been defended in the realm of epistemology by Linda Zagzebski and Arnon Keren, often relying on Joseph Raz's influential reliabilistic argument (see Keren 2014; Raz 1988, 2009; Zagzebski 2012, 2020).

Nevertheless, *preemptionism* has encountered substantial criticism, both through counterexamples and the identification of theoretical shortcomings (cf. Jäger 2016; and Wright 2016). To address these issues, Constantin and Grundmann (2020) proposed narrowing the definition of *epistemic authority* to instances of epistemic superiority. They contend that within this refined scope, *preemptionism* can be justified through principles of rational epistemic defeat. Our contribution offers a meticulous examination of these approaches to preemptionism, provides a formal model, and elucidates its main underlying reliabilistic assumption.

To achieve this, we initially define the notion of *preemption* and *epistemic authority* as conceptualised by Zagzebski (2012, 2020) in section 2. Next, we discuss the defeatist account of preemptionism of Constantin and Grundmann (2020) and outline preemptionism's main argument, the argument from accuracy, in section 3. Subsequently, we furnish a Bayesian elaboration in section 4. In section 5, we go in detail through a validation of the argument from ac-

[*][This text is published under the following bibliographical data: Feldbacher-Escamilla, Christian J. (2025). "Preemptionism and Its Reliabilistic Assumption: A Bayesian Model". In: *The Epistemology of Experts: New Essays*. Ed. by Brössel, Peter, Eder, Anna-Maria Asunta, and Grundmann, Thomas. New York: Routledge/Taylor & Francis Group, pp. 57–74. DOI: 10.4324/9781003431459-6. All page numbers of the published text are in square brackets. The final publication is available at https://doi.org/10.4324/9781003431459-6. For more information about the underlying project, please have a look at http://cjf.escamilla.academia.name.]

curacy and highlight its main reliabilistic assumption. We conclude in section 6.

2 Expertise, Epistemic Authority, and Preemption

As Zagzebski notes in her book on *epistemic authority*, epistemologists occasionally use the notion of an *epistemic authority*. They do so mainly when they refer to experts (cf. Zagzebski 2012, p.5). [58] Expertise covers one important facet of authority, although there are also important differences and investigations of conditions for being an expert in a field do not automatically bring about conditions for being an authority regarding that field for someone. Constantin and Grundmann (2020, p.4114, but also Linda Zagzebski in sect.8.1 in her contribution in this volume) have made this difference clear by stressing that oftentimes expertise is only domain-relational whereas being an authority is domain- and subject-relational (note that Martini 2019, frames 'expertise' also as subject-relational, namely as relational with respect to a layperson, however, the crucial difference seems to be also in his account a domain-relational aspect, distinguishing experts from laypeople): someone can be an expert in a field, an authority, e.g. for a layperson, and at the same time also no authority, e.g. for another expert in this field that is even a luminary. So, the difference is that *expertise* is a kind of threshold-notion—gaining *enough* knowledge and reasoning skills in a domain makes one an expert—, whereas *authority* is a relational notion, regardless of whether one passes a threshold or not. However, very often, and particularly always in case when we compare laypeople with experts, also the relational aspect is what matters most for the consideration of how to incorporate expertise in one's thinking and reasoning. So, in fact, what oftentimes matters is that there is an epistemic gap in the sense that an expert (contra a layperson) or an authority is considered to be epistemically superior, which means that it is supposed to have at least as good evidence and can draw inferences from the evidence at least as good as the individuals who are subordinated to it (cf. Constantin and Grundmann 2020, p.4114). So, an authority is believed to outperform a subordinate individual with respect to evidence *and/or* inferential skills.

Now, coming from the practical domain, *superiority* has been linked to what has been called 'preemptive power' (cf. Raz 1988, 2009). For the domain of law, politics, and decision-making in general, Raz has proposed an influential theory of authority that sees at its core the feature of having normative power, generating *preemptive* reasons for others based on which they rationally make their decisions (cf. Raz 1988, p.47, 2009, pp.385f). The underlying idea is that if an agent or institution a is an authority for an agent or subject s regarding a domain or realm of power p, a can in principle provide reasons for s regarding decisions in p such that the reasons regarding p provided by a can replace that of s and s can in principle *rationally* base her decisions within p on these reasons. Keren (2007) and Zagzebski (2012, 2020) suggested to expand this idea also to the epistemic realm and by this characterise the manifold notion of an

epistemic authority. Roughly put, the idea is to consider the epistemic authority of a over s regarding a specific domain of beliefs and propositions p to consist of s trusting a's "way in which [she] gets [her] belief more than the way in which [s] would get the belief. [59] In cases of this kind the conscientious thing to do is to let [a] stand in for [s] in [s'] attempt to get the truth in that domain [p] and to adopt [her] belief" (cf. Zagzebski 2012, p.105).

The notion of *preemption* is linked also to other features of authorities, namely *content independency*, *subject dependency*, and *normal justifyability*. Content independency demands that if a is an authority for s regarding a domain or realm of power p, then a has normative power over s regarding all reasons relevant for p. So, e.g., if a legislative board is an authority for an individual driver regarding the road traffic act and provides a reason for driving on the left side by ordering so, it could have—with the same authoritative force—also ordered driving on the right side. An individual driver disregarding her reasons for making a decision on riding on a particular side in favour of the authority's reasons to ride on the left side via an order provided by the road traffic act, would be also demanded to disregard her reasons and base her decision to ride on the right side, if the legislative board had ordered to do so via the road traffic act. In the epistemic realm, content independency demands of s, e.g., that if s bases her belief in a proposition p on the reason that authority a believes p, then s would be also supposed to base her belief in $\neg p$ on the reason that authority a believes $\neg p$. In an explication of the *preemptive view* as outlined below, content independency becomes relevant in the sense that an authority's sphere of action ranges not only over single propositions of a domain, but a whole algebra.

A general worry is that deference to authorities may be incompatible with self-reliance and autonomy, values that were put forward especially in epistemological matters during the Enlightenment. This opposing trend is also one of the reasons why speaking of authorities might be like a red rag to a bull in the epistemic realm (cf. Zagzebski 2012, chpt.1). For this reason, the notion of an *epistemic authority* is usually supplemented by a condition that is "sensitive" to the subject that finds oneself subordinated under an authority, the subject-dependency condition—cf. the 'Dependence Thesis' in (Raz 1988, sect.3.2): a is an epistemic authority for s regarding p only if s believes that a "does what [s] would do if [s] were more conscientious or better [at] getting the truth" regarding p (cf. Zagzebski 2012, p.109). The idea is that by this, s is not blindly deferring to an authority, but reflectively and rationally insofar as she has a belief about the epistemic superiority of a in mind. In our later explication of the *preemptive view*, subject dependency is partly covered by the subject-relatedness of credences.

Note that the condition of subject-dependency only links an epistemic authority a s-supposedly to the truth of p. This might provide some internal justification for s to replace her reasons for or against p by that provided by a. [60] However, if s' beliefs about a are not correct or not rationally tracking a's reliability with respect to p, then there is no external justification for demanding s to let a preempt her reasons. In order to guarantee epistemic justification, the

condition of normal justifyability enters the scenery (cf. the Justification Thesis 1 in Zagzebski 2012, p.110): "The authority of [a]'s belief for [s] is justified by [s'] conscientious judgment that [s] is more likely to form a true belief and avoid a false belief if [s believes] what the authority believes than if [s tries] to figure out what to believe [by herself]." If 'justification' is used internally here, *normal justifyability* only provides internal reasons. If it is used externally here, it provides also external reasons. In our debate, we won't decide for one or the other approach. Rather, we want to leave open whether preemptionism should be justified internally or externally. In case of an internal justification, what follows is always to be read under the pre-text of s supposing or considering something, e.g., s supposing or considering a to be more reliable etc. In case of an external justification, what follows it to be read without such a pre-text, so, e.g., a is more reliable than s *simpliciter*.

Note that Zagzebski discusses also a constraint for justification that links the reasons of a not to the truth of p, but to conscientious self-reflection. By this, justification shifts more towards coherence (with respect to p). However, both constraints can fall apart (cf. Wright 2016, sect.2), and for reasons of simplicity we focus here on "normal justifyability" in terms of truth-conduciveness.

Considering only a single proposition p of a particular type about the domain in question, and considering only a set-up with two agents, we can summarise the conditions and features of an epistemic authority as follows:

Epistemic Authority
Some a, believing a proposition p (schematically: $Bel^a(p)$), is an epistemic authority for subject s regarding p iff

1. *Epistemic Superiority Condition*: *Given the beliefs of s*: a is supposedly epistemically superior to s with respect to p in the sense that:
   - i a has supposedly at least the same domain-specific evidence concerning p as s does or even epistemically better domain-specific evidence, and
   - ii a is supposedly at least as good or even better in the sense of being more reliable in drawing correct inferences from the evidence concerning p than s is.

   We assume that at least for one of these conditions a is not only supposedly at least as good but supposedly strictly better than s.
2. *Preemption Condition*: $Bel^a(p)$ is a preemptive reason for s to believe in p, i.e. for $Bel^s(p)$. This means in particular that $Bel^a(p)$ is supposed to screen off s' reasons or that $Bel^a(p)$ replaces s' reasons. [61] According to the latter approach of replacement, the following conditions need to be amended:
   - *Content independency*: If it were the case that $Bel^a(\neg p)$, then this would be also a preemptive reason for $Bel^s(\neg p)$

- *Subject dependency*: s believes that she would also believe in p if she were better at getting at the truth of p, as she believes a to be.
- *Normal justifyability*: s is epistemically justified in this belief about a with respect to p.

Note that in our formulation these conditions hold only supposedly—given the beliefs of s. Again, this expresses the possibility of considering only the purely internalist question under which conditions it is rational *for subject* s to consider a as an epistemic authority. If one takes out 'supposedly', then one ends up with an externalist version. To put it in other words: As characterised here, we are wondering whether, given s rationally *believes* that a is epistemically superior and provides preemptive reasons, it is rational for s to consider a as an epistemic authority. However, it is also possible to have a structurally analogous discussion for the case that a is, in fact, epistemically superior and provides preemptive reasons, and whether for this case s should consider a as an epistemic authority.

Also, we should mention that the conditions on epistemic superiority put forward here are very strong, particularly the condition on the superiority with respect to evidence (1.i). That the authority or expert a has all the evidence of the subordinated subject s or at least defeating evidence, is a condition too strong to be met in many real-world cases. However, our restriction is along the same lines as Constantin and Grundmann (cf. the possibility of a having defeaters with respect to s' evidence in the next section; for a discussion of the relevance of domain-specificity cf. their 2020, sect.3) and aims at improving the notion of Zagzebski (2012) by putting forward strong conditions for its application. Regarding the preemption condition (2), we find the "screening off"-condition in (Constantin and Grundmann 2020, p.4111) and the "replacement"-condition in (Zagzebski 2012, p.102).

Regarding the conditions amending the preemption condition, subject dependency and normal justifyability provide greater initial plausibility for the preemption condition: If one *rationally thinks* that an epistemic authority acts the same way as one *would act* if one had more evidence and inferential skills, why not let the authority's belief stand in for one's own? Furthermore, subject dependency also serves as a defeater inasmuch it allows one to rule out "crazy" beliefs of seemingly authorities. If, e.g., a physician believes you should take 4,000 pills in order to cure a disease, you might easily conclude that even with more evidence and further inferential training you would never end up with such a belief [62] and by this you rule out the physician as an epistemic authority (cf. Zagzebski 2012, p.116; note that Zagzebski 2016, p.188, also discusses the switching of epistemic attitudes of an epistemic authority concerning one and the same proposition as a possible defeater, which can be accounted for if the preemption condition is read externally, i.e. if s only preempts if a is *truly* more reliable, because in case of such a switching regarding p, a is no longer truly more reliable with respect to p). We consider the amended conditions to be relevant for providing some initial plausibility of the preemption condition.

We think that its main justification comes, however, from a different source. In the next section, we elaborate more on the exact link between the two conditions (1 on *epistemic superiority* and 2 on *preemption*) and by this want to work out the details of this justification.

3 Defeatist Preemptionism

In the previous section, we spoke about linking the notion of an *epistemic authority* in the sense of an epistemic superiority to the notion of *preemption*. Such a link can be of two sorts: it can be a thesis about supposedly epistemically superior agents providing also preemptive reasons; or it can be about the *preemption condition* being even a definitional feature of the notion of an *epistemic authority*. Both claims seem to be feasibly attributed to different accounts in the literature. E.g., Raz speaks about the 'pre-emption thesis' and argues for its truth (cf. Raz 1988, p.47, 2009, p.391). Zagzebski, on the other hand, although she also refers to it with 'principle' and 'thesis', seems to consider it a definitional feature: the index of her book lists it within 'authority: definitions of', at some other occasion (at a book-symposium at the Pacific APA; cf. Wright 2016, p.557, fn 3) she suggested to consider the "preemption thesis" a definition of 'authority', and also her argumentation in favour of preemptionism aims at defending the "existence" of epistemic authority (cf. Zagzebski 2012, p.2).

Technically, the difference is about how the listed features of an epistemic authority at the end of the previous section (i.e. 1 and 2) are employed. One can consider 1 *and* 2 as a definiens. Or one can go for only 1 as the definiens of the notion of an *epistemic authority* and then try to show that 2 is a consequence of it (given a broader framework of epistemic rationality). In principle, we think that one does not need to take a particular stance here. Accepting criticism of accounts of epistemic authority (cf., e.g. Jäger 2016; Wright 2016) can be oftentimes equivalently taken into account by strengthening *definitional* preemption conditions (having 1 and 2 as a definiens) as well as by adding proviso-conditions to a preemption *thesis* (1 implies 2). However, there is an advantage of treating the role of the preemption condition as that of being definitionally independent of the notion of an *epistemic authority*. [63] Keeping it that way allows modifications due to the discussion of counterexamples not triggering any changes in the conceptual core (i.e. in a definition with only 1) but only in applications of the concept (i.e. regarding a possible implication of 2). Such a strategy is, e.g., suggested in the account of "defeatist preemptionism" of Constantin and Grundmann (2020). They argue that supposedly *epistemic superiority* should make up for the notion of an *epistemic authority* but that also a particular form of preemptionism, namely *defeatist preemptionism*, follows from *epistemic superiority* and by this can be also directly attributed to *epistemic authority*.

In order to show the latter, they suggest to use as a rational background theory that of *source sensitive defeaters*. Source sensitive defeaters bring it about that "it is generally irrational to rely on evidence obtained from [a] source the de-

feater calls into question" (cf. Constantin and Grundmann 2020, p.4120). Typical source sensitive defeaters are, e.g., what one finds in the literature as so-called "undercutting" defeaters (cf. Pollock and Cruz 1999) and "higher-order" defeaters (cf. Christensen 2010). Undercutting defeaters call the veracity of evidence in question (e.g. knowing that it is red light and not daylight that is shining on objects when assessing the colour of objects) whereas higher-order defeaters call the processing from evidence towards a (hypo)thesis in question (e.g. knowing that someone has cognitive biases). In both cases, one should no longer use one's own evidence and reasoning. Now, Constantin and Grundmann (2020, p.4122) suggest that according to their understanding of preemptionism in a defeatist way, "the preemptive force of authoritative reasons is explained as a special case of [source sensitive] defeat". The main idea is that the belief of an authority is nothing else than a source sensitive defeater for a subordinated subject and since source sensitive defeaters render use of one's own evidence (undercutting) and reasoning (higher-order) irrational, "learning about the belief of an authority renders further use of one's own evidence in assessing the content of that belief irrational" (Constantin and Grundmann 2020, p.4117). The reason they provide for this is as follows:

> "[I]f you learn that the authority has formed her attitude while taking into account all of your relevant evidence, then you also have good reasons to believe that the authority's credence is adequate based on evidence that includes yours and, at the same time, that your own credence would be inadequate insofar as it would differ from the authority's. Moreover, relying on any evidence in addition to the authoritative reason for the assessment of p would lead to a deviation from the credence the authority assigns to the proposition. Hence, any reliance on evidence other than the authoritative reason itself would render your credence less likely to be adequate." (cf. p.4120)

[64] Such reasoning for preemptionism through *inferential superiority* was also stressed by Zagzebski and Raz by help of the so-called *argument from accuracy*. The argument states that if one does not disregard one's epistemic attitudes in favour of that of an authority, one falls short of approaching the authority's reliability. The reason is that not doing so waters down one's performance via one's own epistemically inferior influence (cf. Raz 1988, p.68; although all of the accounts mentioned here use reliability reasoning, there are differences with respect to the aim of this reasoning; whereas Zagzebski uses it to argue for preempting all reasons of s, Raz only uses it to argue for preempting all contra-reasons of s; this difference was spelled out by Dormandy 2018, and has led her to establish another account of epistemic authority, namely a *proper basing account*). Here is our reconstruction of the general form of this argumentation:

1. In case of an epistemic authority, i.e. supposedly evidential and inferential superiority of a over s regarding p, a is supposed to be better in tracking the truth of p than s, i.e., regarding p, the reliability of a is supposed to be higher than that of s.

2. If s preemptively takes over the credences of a regarding p, then s' reliability approaches that of a.

3. If s only partly takes over the credences of a regarding p (e.g. by averaging over her credence and that of a), then s' reliability falls short of that of a.

4. Hence, in case of a supposedly epistemic authority, reliabilism is in favour of preemptionism.

With premiss 1, we assume that for the domain in question a is supposedly epistemically superior to s. Premiss 2 simply states that in case that s in fact preempts in favour of a, i.e. s mimicing a, the dynamics is such that also the reliabilities will converge. Premiss 3 states that any alternative way to deal with the evidence about the epistemic attitude of an authority a will lead to a less reliable state for s. Given these premisses, we can in fact conclude that, 4, preemptionism fares better in terms of reliabilism than not to preempt but rather only to partially take an authority's credence into account. So, the argument is valid. But is it also sound? Premiss 1 is more or less a stipulated truth. Premiss 2 is more or less a triviality (same epistemic states lead to same reliabilities, if we measure reliabilities on the basis of epistemic states). Premiss 3, however, is a strong claim about any alternative's inferior performance and has to be investigated further. In order to do so, we provide a detailed (Bayesian) model of preemptionism in the next section. Subsequently, we study which notion of reliability is involved in a justification of premiss 3. [65]

4 A Bayesian Model

We can explicate (defeatist) preemptionism within the framework of Bayesian epistemology. There we have a quantitative framework that takes credences as the units of evaluation and change. The easiest set-up contains as a structure the credences of individual s and the epistemic authority a. The structure of the problem is enriched by adding further subjects and authorities (in the latter case, e.g., there might arise a problem of conflicting authorities), however, since we aim at introducing a simple Bayesian model, we stick to the simple setting of two epistemic agents only. This suffices in order to elucidate the justification of the premisses of the argument from accuracy for preemptionism. So, the formal set-up contains a credence function for s, Cr^s, and a credence function for a, Cr^a. Both individuals base their credence function on evidence e^s and e^a respectively.

The problem of how to incorporate evidence within the Bayesian framework concerns the question of how to update on new evidence. Classical update rules such as Bayesian updating for updating on certain evidence and Jeffrey conditionalisation (for updating also on uncertain evidence, cf. Jeffrey 1983) tackle this problem. We will focus on Bayesian updating here. All things

said hold, however, also for the more general version of Jeffrey conditionalisation too. Bayesian updating asks for equating the unconditional credence of some proposition p after receiving some evidence e, i.e. the posterior credence of p, with the conditional credence of p in the light of e, i.e. the prior credence of p given e. If we take as transition point $t \rightarrow t'$ the time where some agent received evidence e and updates on it from her prior credence Cr_t to her posterior credence $Cr_{t'}$, Bayesian updating demands of s and a:

$$\underbrace{Cr^s_{t'}(p) = Cr^s_t(p|e^s)}_{\text{update on evidence of } s} \qquad \underbrace{Cr^a_{t'}(p) = Cr^a_t(p|e^a)}_{\text{update on evidence of } a}$$

This update rule and model contains all needed ingredients for spelling out epistemic superiority of a with respect to s concerning the proposition or domain p: First, e^a has to be at least as good or better than e^s. In the simplest case, this means that e^a is logically at least as strong as e^s, i.e. $e^a \vdash e^s$. Another case of evidential superiority concerns e^a defeating e^s, which means, as we have seen in the previous section, that in the light of e^a, it is irrational to rely on e^s. In the Bayesian framework, this means that it is irrational to update on e^s given e^a. Now, soon we will see that we do not need to exactly model e^a defeating e^s here because our formulation of preemptionism will overwrite any impact of e^s on s' posterior credence (an exact explication of defeat would demand us to formulate statements about higher-order updating; however, for validating the argument from accuracy, we do not need to zoom into that much detail). So much for evidential superiority. [66] What about inferential superiority? It means that a is better in the sense of more reliable and more accurate in inferring the truth of p from some evidence e than s is. If we take v to be an evaluation function providing us information about whether p is true ($v(p) = 1$) or false ($v(p) = 0$), then we can define the accuracy of some credence as, e.g., the inverse (within the unit interval) of the distance between $v(p)$ and the credence $Cr(p)$. Without restriction of generality, we can take as a general proxy for our discussion the quite common distance measure of the squared difference of the evaluation and the credence. If p is about a probabilistic event, a similar apparatus can be employed for defining and measuring accuracy. In that case, v is not a 0-1-function but a probability function. The accuracies of s' and a's credences (A^s, A^a) are:

$$\underbrace{A^s(p) = 1 - (v(p) - Cr^s(p))^2}_{\text{accuracy of } s} \qquad \underbrace{A^a(p) = 1 - (v(p) - Cr^a(p))^2}_{\text{accuracy of } a}$$

Epistemic superiority in terms of accuracy amounts to the accuracy of a superseding that of s, i.e. $1 - (v(p) - Cr^a(p))^2 > 1 - (v(p) - Cr^s(p))^2$.

Reliability is usually attributed to processes and is about the truth-aptness or veracity of a process (cf. Audi 2011, chpt.10). However, it is also not uncommon to speak of the reliability of an epistemic agent, as, e.g. when we speak about Ann's reliability in predicting the correct weather. We follow the latter mode of speaking but link it to the former by assuming that all the relevant

credences of an individual have been formed by the same process and that, therefore, we can evaluate the process by just looking at the accuracy of the prediction of the individual. Also, since we are in the quantitative realm of credences and not only in the qualitative realm of beliefs, disbeliefs, and abstaining of beliefs, we fine-grain the veritistic consideration by, in fact, measuring accuracy instead of simply counting qualitatively correct instances. In more detail, reliability in our set-up amounts to averaging over accuracy among a set of propositions $\{p_1, \ldots, p_n\}$ of the same type as proposition p:

$$\underbrace{\frac{\sum_{i=1}^{n} A^s(p_i)}{n}}_{\text{reliability of } s} \qquad \underbrace{\frac{\sum_{i=1}^{n} A^a(p_i)}{n}}_{\text{reliability of } a}$$

One can also take into account evidence e in such a reliability measure, if one conditionalises the credences on e. If we combine both forms of superiority, evidential and inferential, we can say that a is an epistemic authority for s with respect to p if, e.g. [67], $e^a \vdash e^s$ and the reliability of a ($\sum_{i=1}^{n}(1 - (v(p_i) - Cr^a(p_i|e^a))^2)/n$) is greater or equal to the reliability of s ($\sum_{i=1}^{n}(1 - (v(p_i) - Cr^s(p_i|e^s))^2)/n$).

So much for the problem of how to define or identify an epistemic authority in the Bayesian framework. How about expanding this framework to the problem of how to deal with an epistemic authority? So, how should a subject s update her credences in the light of her getting to know the credences of an epistemic authority a? In order to not have to think about the individual evidence for a moment, let us assume that both agents have updated already on their evidence at the transition point $t \to t'$, so we get the credences $Cr^s_{t'}(p)$ and $Cr^a_{t'}(p)$. If we take now $t' \to t''$ to be the transition point where s learns about the credences of a regarding a proposition p, the problem of epistemic authority concerns the question of how to update her prior credences $Cr^s_{t'}$ to posterior credences $Cr^s_{t''}$. 'Prior' means here prior to getting to know an authority's credences, and 'posterior' means here after getting to know them. So, $Cr^a_{t'}(p) = x$ serves as new evidence for s. Now, although the literature on preemptionism and "screening off" or "unhinging" or "basing" of reasons is oftentimes about the detailed process of preemption (cf. Constantin and Grundmann 2020; Dormandy 2018), our model focusses mainly on the result of such screening off, unhinging or basing. According to preemptionism, the result would be simply s' deference of her epistemic assessment to that of a. This can be formulated as follows:

> **Preemptionism/Deference**
> If a is an epistemic authority for subject s regarding p with credence x at the transition point $t' \to t''$, then s takes over a's credence on p. So, it holds:
> *Credence update*: $Cr^s_{t''}(p) = Cr^s_{t'}(p|Cr^a_{t'}(p) = x) = x$

Note that preemptionism or deference really adds something to Bayesian updating: If s learns at the transition point $t' \to t''$ that $Cr^a_{t'}(p) = x$, the first

equation ($Cr^s_{t''}(p) = Cr^s_{t'}(p|Cr^a_{t'}(p) = x)$) concerns ordinary Bayesian updating. However, the second equation ($Cr^s_{t'}(p|Cr^a_{t'}(p) = x) = x$) goes over and above Bayesian updating and expresses the core of preemptionism deference in this model. Note that we consider it still a *Bayesian* model because Bayesian updating still plays the central role in the model—*Bayesian* in the sense that the posterior credence results from conditional prior credences, even if the result were to be modified or generalised Jeffrey style. That in case of an epistemic authority the update is further constrained does, so it seems, not speak against Bayesianism as such but only against a purely subjectivist interpretation of Bayesianism. [68]

As we have stated above, if e^a is a(n undercutting) defeater of e^s, s is supposed to no longer update on e^s. In this model, this is achieved basically by letting the update on e^s (from t to t') run void in the overall process from t to t'' because s' update from t' to t'' overwrites any previous impact of e^s (from t to t').

Now, Credence update of preemptionism can be reformulated as follows:

$$Cr^s_{t''}(p) = 0 \cdot Cr^s_{t'}(p) + 1 \cdot Cr^a_{t'}(p)$$

This allows us to interpret the update of s after learning the credence of a regarding p as a weighted average of the prior credences of a and s, where the prior credence of a gets full and the prior credence of s gets zero weight. The relevant competitor of the preemptive view is the so-called *total evidence view*, which argues that the weighting should not (always) be 0:1, but cover the full range of [0-1]:[0-1] (cf. Kelly 2011).

Note that this Bayesian model of how to deal with getting to know the epistemic attitude of an epistemic authority does not imply that the epistemic authority is infallible or supposed to be infallible. The latter would mean that a has maximal reliability with respect to p, i.e. full accuracy with respect to any p_i of type p: $A^a(p_i) = 1$, or at least supposedly so. This is, however, not presupposed in this model and, as we will see in our discussion below, also not necessary to argue for the epistemic value of this model. Rather, what is important is that with respect to p the reliability of a is above the reliability of s.

We have the main ingredients now that we need in order to assess the main argument of preemptionism as outlined in section 3: we have explicated the deference of preemptionism and a measure for reliability. Let us see now, how exactly premiss 3 of the argument from accuracy for preemptionism can be justified.

5 The Argument from Accuracy for Preemptionism

Recall the argument from accuracy as presented at the end of section 3: Reliabilism is in favour of preemptionism/deferrence because in case of an epistemic authority, we have a supposedly higher reliability of the authority (pre-

miss 1), the subordinating individual approaching that reliability by preempting/deferring (premiss 2), and the subordinating individual falling short of that reliability if it only partly preempts/deferrs (premiss 3). Now, whereas premisses 1 and 2 clearly hold for the reasons mentioned at the end of section 3, premiss 3 does, strictly speaking, not hold. If we are really only spotting reliability without any modification, then preemptionism is not providing any guarantees about alternatives falling short. [69] To see this, think about the simple measure for the reliability of an epistemic agent that takes the average of the inverse (within the unit interval) of the squared distances of her credences in a series of propositions $p_1, \ldots, p_n$, which are all supposed to be of the same type as p, from the true value, provided via the evaluation function v from the previous section. Now, as the following quite arbitrary example illustrates, sometimes to not completely defer might fare better than preemptionism/deferrence in terms of reliability. Let us assume that the credences Cr^s and Cr^a, without any higher-order considerations, are as follows:

	$\overbrace{\quad\quad\quad\quad}^{t'}$				$\overbrace{\quad\quad\quad\quad\quad\quad\quad\quad}^{t''}$				$\overbrace{\quad\quad\quad\quad}^{t'''}$	
	p_1	p_2	$\ldots$	p_n	p_{n+1}	p_{n+2}	$\ldots$	p_{n+m}	p_{n+m+1}	$\ldots$
Cr^s	0.0	0.0	$\ldots$	0.0	0.0	0.0	$\ldots$	0.0	0.0	$\ldots$
Cr^a	1.0	1.0	$\ldots$	1.0	1.0	1.0	$\ldots$	1.0	1.0	$\ldots$
v	1	1	$\ldots$	1	1	1	$\ldots$	1	0	$\ldots$

Table: Counterexample to premise 3

The true outcomes are given by v—all propositions up to p_{n+m} turn out to be true, all the others are false. Now, assume $p_1, p_2, \ldots, p_n$ to be cases that are relevant for s to establish that a is an epistemic authority for s (in the scheme of our Bayesian model in the previous section, that information is gained before the transition point $t' \rightarrow t''$). Since a's credences match the truth perfectly, its reliability is 1.0 (in 100% of the cases a's credences were in full accordance with the truth). Conversely for s: its reliability is 0.0. Let us assume that this is all transparent to s, so, for s, a is supposedly an epistemic authority with regards to propositions of type p. And, also *in fact*, a is epistemically superior to s regarding propositions of the same type as p within the time frame t'. Now, above we have the initial credences without any higher-order considerations, simplified speaking without any updating. Let us see what happens if s updates. According to preemptionism, s has to let a preempt her epistemic attitude by taking on the credences of a regarding $p_{n+1}, \ldots, p_{n+m}$. If m is sufficiently greater than n, s' reliability will approach (in accordance with premiss 2) that of a, namely 1.0, although s will be still slightly inferior with respect to a. Given that all of this is epistemically transparent to s, a is still an epistemic authority for s within the period t''. Now, assume that at some point in time $n + m + 1$ (i.e. the transition point $t'' \rightarrow t'''$), the v-value changes in favour of s' initial credences. If the series is long enough and significantly exceeds $n + m$, we can see that s could have done significantly better than approaching the reliability of a if she were not deferring to a. [70] If s would not have let a preempt her epistemic

attitude, but would have weighted her credence fully and disregarded that of a—so if s performed a different strategy than suggested by preemptionism—, then, at some point in time, s would have overtaken a in terms of reliability by not following preemptionism. Note that due to the recognised inferiority with respect to $p_1, \ldots, p_n$, a remains epistemically superior compared to s, if s lets her credence be preempted by a. So, it can easily happen that preemptionism in case of even truly supposedly epistemic inferiority is outperformed by an alternative approach. That s defers to a is what makes it in fact never superseding a's performance. This is a case where, if s were to only partly or not at all taking over the credences of a regarding p, s' reliability would not fall short of that of a. Hence, premiss 3 is, strictly speaking, not true. Lackey (2018, p.238) discusses other "alternative policies that would have even better epistemic results [such as]: follow the advice of an authority, except [...] when one knows that the authority is wrong".

However, to accept any technical possibility as counterargument against an epistemic account would render the epistemic enterprise of justifying epistemic accounts impossible—given such a standard, all epistemic accounts would be virtually on a par. A more moderate and more discriminative standard is to test whether preemptionism is, on average, faring better than any deviation from it. This means, however, to modify the notion of *reliability* in the argument under consideration and particularly in premiss 3: it is not simply about reliability, but about *expected* reliability. Zagzebski argues for such a claim by the help of reference to problems of *probability matching* (cf. Zagzebski 2012, p.115). The problem of *probability matching* is as follows (cf. Vulkan 2000, sect.2): Given two mutually exclusive options p_r (right) and p_l (left) that are randomly distributed with fixed probabilities $Pr(p_r)$ and $Pr(p_l)$, what is the right strategy to make a decision for one of the options? As empirical studies show, humans tend to perform the so-called strategy of *probability matching*. According to this strategy, the frequency of one's decisions for an option matches the probability of that option. So, if, e.g., p_r shows up 75% of the time and p_l only 25% of the time (i.e. $Pr(p_r) = 0.75$, $Pr(p_l) = 0.25$), then humans, when asked which option to choose, tend to opt for p_r also 75% of the time and for p_l 25% of the time (cf. Gallistel 1993, chpt.11). Non-human animals like rats act differently: They perform a *take-the-most-frequent* strategy that favours exclusively that option which has a higher probability. So, after some phase of learning the probabilities of the example above, they opt for p_r exclusively. What is the rationale of both strategies? It is easy to demonstrate that the expected utility of the *take-the-most-frequent* strategy is maximal (cf. Vulkan 2000, sect.2): If we calculate the expected utility of making the correct prediction for an agent s having credences Cr^s as follows:

$$Cr^s(p_r) \cdot Pr(p_r) \cdot u(p_r) + Cr^s(p_l) \cdot Pr(p_l) \cdot u(p_l)$$

[71] If we furthermore assume that p_r and p_l are jointly exhaustive and mutually exclusive defining the probability space; and if we finally assume that the utilities of correctly predicting p_r and p_l are equal ($u(p_r) = u(p_l)$), then it turns out that, e.g., having $Cr^s(p_r) = 1.0$ in case $Pr(p_r) \geq Pr(p_l)$ maximises

the expected utilities for s: Due to these assumptions, the expected utility for s making a right prediction is proportional to:

$$Cr^s(p_r) \cdot (2 \cdot Pr(p_r) - 1) + 1 - Pr(p_r)$$

So, if $Pr(p_r) \geq Pr(p_l)$, then $Cr^s(p_r) \cdot [0.0, 1.0] + [0.0, 0.5]$ (the possible values under this assumption) is maximised by $Cr^s(p_r) = 1.0$. And if $Pr(p_r) < Pr(p_l)$, then $Cr^s(p_r) \cdot [\text{-}1.0, 0.0] + [0.5, 1.0]$ (the possible values under this assumption) is maximised by $Cr^s(p_r) = 0.0$. On average, the highest, i.e. the maximal expected, utility is gained by the *take-the-most-frequent* strategy: In the example above it will decide in 75% of the cases correctly. *Probability matching* is on average below, although, of course, in single instances it might perform better.

This case can be applied to the discussion of preemptionism as follows: If a is an epistemic authority for s regarding p, then a is supposedly more reliable regarding p than s is. If s follows preemptionism, then s will approach the reliability of a regarding p: According to preemptionism/deference, s has to update her credence $Cr^s(p)$ to that of $Cr^a(p)$ via $Cr^s_{t''}(p) = Cr^s_{t'}(p|Cr^a_{t'}(p) = x) = x$. What is more, it even will maximise her "expected" reliability, i.e. her reliability on average with respect to that of a. This implies that if s deviates from preemptionism/deference, s will, on average, fall short of achieving the reliability of a, despite the possibility of outperforming a in terms of reliability in individual cases. So, examples as the one provided above where s outperforms a by deviating from preemptionism/deference are not representative for average cases. And this provides a justification for practicing preemptionism instead of its alternatives. In terms of *expected* reliability as an epistemic end, premiss 3 is, in fact, true.

So, this shows, that the argument from accuracy is, in fact, an argument about *expected* accuracy or reliability:

1. In case of an epistemic authority, i.e. supposedly evidential and inferential superiority of a over s regarding p, a is supposed to be better in tracking the truth of p than s, i.e., regarding p, the *expected* reliability of a is supposed to be higher than that of s.

2. If s preemptively takes over the credences of a regarding p, then s' *expected* reliability approaches that of a.

3. If s only partly takes over the credences of a regarding p, then s' *expected* reliability falls short of that of a.

4. Hence, in case of an epistemic authority, *expectation*-reliabilism is in favour of preemptionism.

[72] Now, assuming expectation-reliabilism is not uncommon in the field. In his discussion of epistemic acceptance practices for testimony, Goldman (1999, p.110) writes, e.g.:

> "Is there any acceptance practice that is optimal in all reporting environments, in other words, better in each reporting environment

> than every other acceptance practice would be? As in game theory, the answer appears to be 'no.' [...] Here is a more modest project for the epistemology of testimonial acceptance. [...] A good practice is one that produces veritistic improvements on *average*, over a range of actual and possible applications.

So, also here we find a reference to expectation values. The main point we want to make is that it is important to highlight the exact assumptions about reliability made in the discourse. If it is expected reliability we want to optimise, then preemptionism/deference is justifyably the way to go in case of facing an epistemic authority. If other epistemic ends are aimed at, as, e.g., some kind of avoidance of a problematic switching of one's epistemic attitudes (for details cf., e.g., the switching problem discussed in Jäger 2016), then preemptionism/deference might run counter these epistemic ends. Another case in point against solely considering expected reliability as a benchmark for validating an account of epistemic authority could be made, e.g., on the basis of discussions of epistemic bribery (cf. the discussion of epistemic bribery in Greaves 2013; and for instrumental and teleological aspects Berker 2013; Firth 1981, for a discussion of the role of so-called "zetetic" reasoning in the context of epistemic authorities and particularly experts cf. also the contribution of Dellsén and Linnebo in this volume): If one thinks that an account on epistemic authority should aim at an intrinsic justification and not on an instrumental/teleological/zetetic one, relying on expectation values seems inept because such a reliance guarantees only approaching a better performance on average, but not for the individual instances. In relying on expectation values, the performance in the individual cases has to be considered as instrumental for the performance of an overall set of cases. However, this contribution did not mainly aim at problematising the preemptive account to epistemic authorities. Rather, we aimed to model the account and explicating the underlying standard for its evaluation. The examples discussed here indicate that there are cases for which this standard is inept.

6 Conclusion

We have discussed the notions of *expertise* and *epistemic authority* and their relation to preemption and deference in epistemology, building on the work of (Zagzebski 2012, 2020; Raz 1988, 2009; Constantin and Grundmann 2020). [73] We have outlined a comprehensive theoretical framework to understand when laypersons should defer to experts or authorities in a domain, emphasizing the relational nature of authority which contrasts with the more absolute nature of expertise. Regarding the former, we have seen that Constantin and Grundmann (2020) contend that the concept of an *epistemic authority* can be limited to instances of epistemic superiority, especially when framed within the theories of source sensitive defeat and reliabilism, thus providing a rationale for *preemptionism*. In our discussion, we have pinpointed that this justification relies on a

specific interpretation of *reliabilism*, which we identify as *expectation-reliabilism*. Given the maximisation of expected reliability as an epistemic end, preemptionism/deference is in fact justified by the argument from (*expected*) accuracy. However, given other epistemic ends, preemptionism/deference might run counter these ends.

Acknowledgements

For valuable discussion and feedback on an earlier version of this contribution, I would like to thank, amongst others, the editors of this volume (Peter Brössel, Anna-Maria Asunta Eder, Thomas Grundmann) as well as Christoph Jäger, Arnon Keren, and Dunja Šešelja.

References

Audi, Robert (2011). *Epistemology. A Contemporary Introduction to the Theory of Knowledge*. Third Edition. New York: Routledge.

Berker, Selim (2013). "Epistemic Teleology and the Separateness of Propositions". In: *Philosophical Review* 122.3, pp. 337–393. DOI: 10.1215/00318108-2087645.

Christensen, David (2010). "Higher-Order Evidence". English. In: *Philosophy and Phenomenological Research* 81.1, pp. 185–215. URL: http://www.jstor.org/stable/20779554.

Constantin, Jan and Grundmann, Thomas (2020). "Epistemic Authority: Preemption through source sensitive defeat". In: *Synthese* 197, pp. 4109–4130. DOI: 10.1007/s11229-018-01923-x.

Dormandy, Katherine (2018). "Epistemic Authority: Preemption or Proper Basing?" In: *Erkenntnis*. DOI: 10.1007/s10670-017-9913-3.

Firth, Roderick (1981). "Epistemic Merit, Intrinsic and Instrumental". In: *The American Philosophical Association Centennial Series*, pp. 5–18. DOI: 10.5840/APAPA2013111.

Gallistel, Charles R. (1993). *The Organization of Learning*. Cambridge, Massachusetts: MIT Press.

Goldman, Alvin I. (1999). *Knowledge in a Social World*. Oxford: Oxford University Press.

Greaves, Hilary (2013). "Epistemic Decision Theory". In: *Mind* 122.488, pp. 915–952. DOI: 10.1093/mind/fzt090.

Jäger, Christoph (2016). "Epistemic Authority, Preemptive Reasons, and Understanding". In: *Episteme* 13.2, pp. 167–185.

Jeffrey, Richard C. (1983). *The Logic of Decision*. Second Edition. Chicago: The University of Chicago Press.

Kelly, Thomas (2011). "Peer Disagreement and Higher Order Evidence". In: *Social Epistemology. Essential Readings*. Ed. by Goldman, Alvin I. and Whitcomb, Dennis. Oxford: Oxford University Press, pp. 183–217.

Keren, Arnon (2007). "Epistemic Authority, Testimony and the Transmission of Knowledge". In: *Episteme* 4.3, pp. 368–381. DOI: 10.3366/E1742360007000147.

— (2014). "Zagzebski on Authority and Preemption in the Domain of Belief". In: *European Journal for Philosophy of Religion* 6.4, pp. 61–76.

Lackey, Jennifer (2018). "Experts and Peer Disagreement". In: *Knowledge, Belief, and God: New Insights in Religious Epistemology*. New York: Oxford University Press, pp. 228–245. DOI: https://doi.org/10.1093/oso/9780198798705.003.0012.

Martini, Carlo (2019). "The Epistemology of Expertise". In: *The Routledge Handbook of Social Epistemology*. Ed. by Fricker, Miranda, Graham, Peter J., Henderson, David, and Pedersen, Nikolaj J. L. L. New York: Routledge, pp. 115–122.

Pollock, John L. and Cruz, Joseph (1999). *Contemporary Theories of Knowledge*. 2nd Edition. Oxford: Rowman & Littlefield Publishers.

Raz, Joseph (1988). *The Morality of Freedom*. Oxford: Oxford University Press.

— (2009). *Between Authority and Interpretation: On the Theory of Law and Practical Reason*. Oxford University Press.

Vulkan, Nir (2000). "An Economist's Perspective on Probability Matching". In: *Journal of Economic Surveys* 14.1, pp. 101–118. DOI: 10.1111/1467-6419.00106.

Wright, Sarah (2016). "Epistemic Authority, Epistemic Preemption, and the Intellectual Virtues". In: *Episteme* 13.4, pp. 555–570. DOI: 10.1017/epi.2016.31.

Zagzebski, Linda (2012). *Epistemic Authority. A Theory of Trust, Authority, and Autonomy in Belief*. Oxford: Oxford University Press.

— (2016). "Replies to Christoph Jäger and Elizabeth Fricker". In: *Episteme* 13.2, pp. 187–194. DOI: 10.1017/epi.2015.39.

— (2020). *Epistemic Values: Collected Papers in Epistemology*. Oxford: Oxford University Press.