EPISTEMIC AUTHORITY AND CONSCIENTIOUS BELIEF

CHARITY ANDERSON

University of Oxford

In *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*[1], Linda Zagzebski engages in a wide-reaching investigation of underexplored epistemic terrain. The work is timely, given the recent interest of contemporary epistemologists in testimony and trust, and she approaches the topic from a novel angle: the rationality of beliefs taken on the authority of others. This accessibly written book covers an extensive span of topics including trust, testimony, and authority, in addition to the intersection of these notions with the domains of emotion, morality, religion, communities, and disagreement. Given the breadth of topics in the book, I will not be able to address all the ideas worthy of discussion. I will concentrate on a few topics. I will first examine the notions that serve as the foundation for the book: rationality, conscientiousness, and trust. I then raise several questions concerning her account of epistemic authority. Finally, I advance a worry regarding the application of her account to the problem of disagreement.

I. RATIONALITY, TRUST, AND CONSCIENTIOUSNESS

A main thesis of Zagzebski's book is that trust in oneself and in others is rationally inescapable. The self-reflective rationally consistent agent will trust others. Zagzebski takes as her opponent the epistemic egoist – someone who believes that her only way of gaining reasons for belief is to rely on her own faculties. The extreme epistemic egoist trusts only in herself she does not trust others (a less extreme egoist trusts others but only when her own faculties provide her with reasons to trust the person).

---

[1] Linda Trinkaus Zagzebski, *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*, Oxford: Oxford University Press, 2012.

After defending the rationality of trust in both self and others, Zagzebski argues that it is rational to take beliefs on the authority of others, and then extends her argument to include the authority of communities as well as moral and religious authorities.

Rationality is central to Zagzebski's project. To be rational, she suggests is 'to do a better job of what we do in any case – what our faculties do naturally' (2012: 30). Our initial model of rationality, on her picture, is the resolution of the experience of dissonance or conflict between our mental states. Resolving dissonance is often automatic, but at times involves conscious deliberation. For Zagzebski, self-trust is rational because it resolves dissonance. More specifically, a general trust in one's faculties is rational because it resolves the dissonance arising from the epistemic circularity objection to justified belief in the reliability of our faculties. Particular trust in our faculties is rational when we are conscientious. Zagzebski understands epistemic conscientiousness as the 'quality of using our faculties to the best of our ability in order to get the truth' (2012: 48). It is the state or disposition of being careful and doing our best. Conscientiousness comes in degrees and requires self-awareness and self-monitoring. It is essential to her overall account that we trust we will succeed more often when we are conscientious. In this way, conscientiousness is our second model of rationality. As we shall see, this notion plays a key role in Zagzebski's defence of trusting epistemic authorities.

Zagzebski extends the rationality of trust in oneself to trust in others. Although she offers no clear set of conditions for when trust in others is justified, her main idea is that when we conscientiously judge that we can trust someone, we are rational to trust them. Zagzebski argues that if we are rational, we cannot fail to trust others, on pain of inconsistency. Her reasoning is as follows:

> When I am conscientious I will come to believe that other normal, mature humans have the same natural desire for truth and the same general powers and capacities that I have. If I have a general trust in myself and I accept the principle that I should treat like cases alike, I am rationally committed to having a general trust in them also. (2012: 55)

She takes it as given that we are all basically alike in our general abilities, and while she admits there are exceptions to this general similarity, she suggests the exceptions will be very limited. The principle of 'treat like alike' she understands as *a priori*.

I worry that this reasoning will apply in far fewer cases than Zagzebski seems to expect. Although most people may have the same general powers and capacities with respect to telling the time or knowing whether it is raining by looking out the window, in less mundane cases our powers and faculties vary in initial aptitude as well as specific training. If a rational requirement to trust others with respect to mundane facts is all that the argument establishes, this severely limits the extent to which trusting others is rationally inescapable and thereby reduces the effectiveness of the argument to oppose the egoist.

Zagzebski's reasoning leads her to affirm the following: the fact that another person believes p always gives me *prima facie* reason to believe p (2012: 58). At first glance this suggestion may seem suspect, but she goes on to explain that by 'reason' she does not mean a *decisive* reason, but rather one with 'genuine weight'. The fact that another person believes p counts in favour of p for me, though it does not always provide enough weight for me to believe p. Furthermore, the weight it provides can be defeated – in some cases it can be defeated quite easily. Given how weakly the principle is intended, the reasoning strikes me as plausible. If not read in light of this qualification, some of Zagzebski's statements about *prima facie* reasons could be misleading. One would be forgiven for thinking that she holds that one gains a sufficient reason to believe p when one discovers that someone believes p. In fact she doesn't think this. Nevertheless, Zagzebski maintains that it is *often* the case that trusting another person provides one with a decisive reason to believe what they believe.

Zagzebski extends the argument for general trust in others to particular trust in the faculties of others. When I discover that another person conscientiously believes p this provides me with a stronger *prima facie* reason to believe p (stronger than if I merely know the person believes p). It appears that she thinks most people exhibit conscientiousness most of the time and to roughly the same degree. She claims: '[w]hen I am believing conscientiously, I come to believe that many [others] are just as conscientious as I am when I am as conscientious as I can be.' (2012: 57) She further claims that when I am observant, 'I will inevitably have excellent evidence of the conscientiousness of others.' (2012: 61) Here we find a potential dilemma: being conscientious is either pretty easy, such that everyone is basically conscientious most of the time, or it is fairly demanding. If it is easy, it will not be able to play the role in solving the problems Zagzebski claims it solves. Or, at least, the solution will be

rather thin. Trusting our conscientious judgments is Zagzebski's advice for resolving various difficult epistemic situations, such as disagreement with those we trust. If we are generally conscientious all the time, then her responses to such problems will amount to the suggestion that we merely continue doing what we are already doing and generally always do. But if conscientiousness is demanding, the assumption that so many people exhibit it so often strikes me as unlikely. Most people do not continuously engage in self-reflection. We clearly do not use our faculties to the *best* of our ability all or even most of the time. But if conscientiousness is rare, then we won't readily have evidence of others' conscientiousness.

It's worth noting that Zagzebski seems to assume that we will know when we are conscientious – that when we are doing our best it will be evident that we are doing our best. This assumption strikes me as mistaken. Consider a hard-working student, eager to make good grades. He prepares for exams and tries to do his best, but while he often performs well he only sometimes does his best. It is easy to imagine that it is not obvious to him when he is doing his best and when he is not – at times he may be agnostic about his efforts until he receives his scores. Access to our own conscientiousness is further attenuated by our proneness to self-deception. When we fail we often console ourselves with the assurance that we did our best. The desire for this assurance provides a motivation to believe that we are doing our best most of the time. Our beliefs about our own conscientiousness are prime ground for self-deception.

One final issue arises in this connection. On Zagzebski's account, 'our only test that a belief is true is that it survives future conscientious reflection' (2012: 50). I ought to trust others when I predict their beliefs will satisfy my future self-reflection. In this way, being rational requires predicting the future in unrealistic ways. Zagzebski does not elaborate on how we determine which beliefs will satisfy future self-reflection, though she doesn't seem to expect that we will have trouble making such predictions. Given that we are not usually in a position to predict what our future evidence will be, it's hard to see how we could make judgments about future reflection with any confidence or accuracy.

II. EPISTEMIC AUTHORITIES

Zagzebski's defence of trusting epistemic authorities is the centrepiece of the book. An epistemic authority, on her view, is 'someone who does

what I would do if I were more conscientious or better than I am at satisfying the aim of conscientiousness – getting the truth' (2012: 109). The following principle, the Justification Thesis 2 for the Authority of Belief, states Zagzebski's main idea:

> (JAB2): The authority of another person's belief for me is justified by my conscientious judgment that I am more likely to form a belief that survives conscientious self-reflection if I believe what the authority believes than if I try to figure out what to believe myself. (2012: 110)

Trusting an authority differs from merely trusting another person. Trusting another person provides one with a *prima facie* reason to believe *p* (a reason that merely counts in favour of *p*), but authorities provide sufficient reasons to believe *p,* and further, authorities provide *pre-emptive* reasons.

Before I discuss pre-emption, I want to raise one question regarding who counts as an epistemic authority. It's not clear whether an epistemic authority must be more conscientious than me or whether someone could be an epistemic authority merely in virtue of having more evidence than me. Zagzebski's characterization of an epistemic authority suggests that an authority needs to be more conscientious than myself. An authority has 'more of the qualities I trust in myself insofar as I am epistemically conscientious' (2012: 108); she is someone who 'can help me to believe as I would believe myself, given my desires, emotions, and other beliefs … '(2012: 111). Moreover, 'the point of epistemic authority is to help me in believing conscientiously' (2012: 111). In any case, these remarks indicate that an epistemic authority is not a general ideal, but a better version of yourself – when an epistemic authority is more conscientious than you, she must be conscientious in a way that resembles you. It seems I cannot rationally take someone as an epistemic authority if her psychic structure is incongruent with my current desires, emotions, and other beliefs. (If an epistemic authority must be more conscientious than myself, this sets a fairly stringent constraint on who can be an epistemic authority for me. Surely some of my desires and beliefs prevent me from being more conscientious than I am, thus the more conscientious version of me will lack those desires and beliefs though it's unclear how I would recognize which beliefs and desires those are in advance.)

In her account of what is *authoritative* about an epistemic authority, Zagzebski draws on contemporary work in legal philosophy – the writing of Joseph Raz in particular. Although in the practical realm authority is

typically exercised through commands, Zagzebski claims that the right to command is not an essential feature of authority. (She does mention that she thinks we are psychologically able to believe on command, but this supposition is not part of her account.) On Zagzebski's picture, what is essential to authority is that authorities generate *pre-emptive* reasons for a subject to act or to believe something. A pre-emptive reason is a 'reason that replaces other reasons the subject has' (2012: 102). The fact that an authority believes *p* is not simply added to one's other reasons; trusting an authority involves allowing the authority to 'stand in' for oneself. (It's worth noting one way in which beliefs taken on epistemic authority differ from those based on testimony: it can be rational to believe *p* on someone's authority merely when I know that that authority believes *p*. She does not have to tell me *p*; in fact, the authority does not even need to know that I exist.)

One concern Zagzebski raises concerning the idea of pre-emptive reasons is that it's not obvious that it is psychologically possible for me to let an authority's belief that *p* pre-empt my reasons. One might worry that I can't just ignore my other reasons. In defence of pre-emptive reasons, Zagzebski considers an example of a pre-emptive reason to act. She suggests that it is possible to stop at a red light for the reason that it is against the law to drive through a red light, even if one has other reasons for stopping, such as the likelihood of being hit. She states, 'if I stop because the law says to do so, that reason has the status of being my reason for stopping. It can be the reason even though I am quite capable of reciting many reasons for and against stopping.' (2012: 113) It appears that for the fact that the law requires one to stop at a red light to be my reason for stopping – and thus to pre-empt my other reasons – it must be the case that my other reasons do not play a role in motivating me to stop. That the law requires me to stop must be my basis for action. The parallel with belief is as follows: to believe *p* on authority the fact that the authority believes *p* must be my basis for believing.

I am sympathetic to Zagzebski's proposal that we sometimes make one of our reasons our sole basis for action or belief. But I worry that the settings where we adopt pre-emptive reasons – and thus the settings where we believe on authority – will be more limited than Zagzebski seems to envision. There are several reasons for this: first, because pre-emption is not always psychologically possible. Consider again a case of action. Suppose an authority commands me to take my mother to lunch on her birthday – something I was already planning to do because

I enjoy spending time with her and want her to enjoy her birthday. It's not clear that I can take her to lunch solely for the reason that the authority commands it – my other reasons are too strong to fail to motivate me. In the case of reasons for beliefs, the situation seems worse. If I already believe that it is wrong to steal, and have good reasons for so believing, I may not be able to believe solely because an authority believes it. My other reasons will inevitably play a role in my belief. The same will hold for many beliefs for which I have strong evidence.

Second, if pre-emption is essential for believing on authority, as Zagzebski suggests, it is not clear that it is possible to believe *p* on authority in a setting where I don't already have some reasons to believe or disbelieve *p.* Pre-emption seems to require prior possession of reasons – an authority's belief can't replace my reasons if I don't have any reasons to be replaced. (Perhaps we should understand a pre-emptive reason as a reason that replaces your other reasons *if you have any*.) Moreover, Zagzebski claims that when we already have beliefs on a topic, it is less likely we will trust an authority, since it is less likely we will judge that the authority's belief will survive our conscientious reflection. The foregoing considerations suggest that believing on authority will occur primarily in settings where either (1) I have no reasons to believe or disbelieve *p,* or, (2) I have weak reasons for or against *p*.

One final question regards whether pre-emption requires a kind of counterfactual stability. What if I let a reason pre-empt my other reasons but I very easily might not have? Suppose I stop at the red light because the law requires it, but if I had been in a hurry and no one was around I would have gone through the light. Is this fact relevant to the settings where I stop when I am not in a hurry? We can assume that when I am not in a hurry I believe that I am stopping because it is the law. It's not clear whether what I would do in similar cases makes a difference to whether a reason counts as pre-emptive. This issue may reach beyond the scope of Zagzebski's aims in this chapter, but I raise it as an area deserving further exploration.

III. CONSCIENTIOUS TRUST AND DISAGREEMENT

We've seen thus far that the notion of a belief that survives conscientious self-reflection plays a key role in Zagzebski's discussion. This theme is repeated throughout the work and is the foundation of the guidance she offers for navigating various epistemic difficulties, such as what one

ought to do in the face of disagreement. Zagzebski casts the problem of disagreement as an intrapersonal conflict, rather than an interpersonal conflict. On her view, disagreement is a problem that arises within self-trust because trusting myself commits me to trusting others. When I disagree with someone I trust, I experience conflict between my self-trust and my trust in the other person.

The appropriate resolution to disagreement with people I take to be conscientious, on her picture, is to believe in accordance with whatever 'I trust the most when I am thinking in the way I trust most, that is conscientiously' (2012: 214). When we are conscientious, we are doing our best and that is all we can do. Zagzebski provides surprisingly little guidance for dealing with cases of disagreement beyond this. She does suggest that beliefs that are central to one's noetic structure will often have survived a great deal of self-reflection; thus, when disagreement involves one of these core beliefs it is unlikely that one will judge that one ought to trust the other person. This way of thinking seems to condone a 'stick to your guns' mentality for at least many of our strongly held beliefs.

Resolution of disagreement involves prediction of what will survive future conscientious reflection:

> Disagreement with people we conscientiously judge to be conscientious should be handled ... in a way that we conscientiously judge will survive conscientious self-reflection. ... What is relevant for me is what I conscientiously believe, and what I predict will satisfy my future self-reflection, given what I conscientiously predict about myself. (2012: 215)

We have already mentioned one difficulty for this methodology: it requires predicting the future. A further question concerns what you ought to do if you know that you are bad at predicting what will satisfy your future self-reflection. Suppose you have loads of evidence that you are fairly unreliable at making such predictions. (Actually this strikes me as quite plausible, given that we often cannot predict what evidence will be available to us in the future and which beliefs will survive future self-reflection will be highly dependent on my future evidence.) Given what she's said thus far, it seems likely Zagzebki's response would be to do your conscientious best – that's all you can do.

It's a bit surprising that having dismissed the usual characterization of the problem of disagreement as not very interesting and failing to constitute a distinct evidential problem – despite the abundance of recent

puzzlement on the topic in the literature – Zagzebski's central advice for resolving the conflict is to be conscientious and do your best (2012: 211). In the face of disagreement we often don't know the best way to resolve the conflict and we don't always know what the most conscientious response is. In this way, her discussion may strike some as insensitive to the nuances surrounding the problem.

CONCLUSION

Although I agree with several of Zagzebski's main conclusions, the reasoning she uses to arrive at these conclusions I find wanting in the ways I've mentioned here. By taking the egoist as her main opponent, Zagzebski ends up treating several complex issues with too blunt an instrument. Arguably, the notion of conscientiousness does too much work for her account given the insufficient discussion of what exactly it is to be conscientious. Nevertheless, she points us in the direction of topics worthy of discussion and has constructed a broad framework within which to grapple with these questions. Any reader will find something of interest to mull over in this book, and it will easily become required reading for anyone working on the concept of epistemic authority.