

Critical Study: *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief* by Linda Trinkaus Zagzebski

Guy Longworth
University of Warwick

1. Introduction

This is an ambitious book. It aims to “...defend the existence of epistemic authority on grounds that almost all modern philosophers would accept”. (p.2.) Perhaps unsurprisingly, I’m not convinced that it achieves its aims. Nonetheless, it contains much of interest and is worthy of further study. I’ll begin with a brief overview, before considering in more detail the fundamental argument.

Zagzebski’s basic defence of epistemic authority takes up the first five chapters. The operative understanding of epistemic authority isn’t explained until chapter 5, when it’s revealed to be an epistemicized version of Joseph Raz’s account (Raz 1988). (It would have been helpful to have the operative notion of epistemic authority introduced earlier.) On this account, someone figures as an epistemic authority only if the following conditions are met: (i) their believing something provides *content-independent* reasons for believing it myself, so that if they had believed something else instead, that would have been a reason for me to emulate them; (ii) their believing something provides *preemptive* reasons to believe it, reasons that replace, rather than add to, my other reasons; (iii) a *dependency thesis* holds, in that their belief is “formed in a way that I would conscientiously believe is deserving of emulation” (p.109); (iv) a *justification thesis* holds, in that it is my conscientious belief that I’m more likely to believe well if I emulate the authority. (pp.105–113).

Zagzebski argues that the major alleged sources of resistance to epistemic authority in the historical record are equivocal (Chapter 1), that trust in one’s own epistemic faculties is rational (Chapter 2), that trust in one’s own epistemic faculties rationally requires trust in others’ epistemic faculties (Chapter 3), and that trust in one’s emotions is rational (Chapter 4). The argument to trust in others’ epistemic faculties is designed to show that emulating others beliefs can be rational. However, it doesn’t follow that one should accept others’ testimony except insofar as their testimony reflects their beliefs. That we can rationally trust others’ testimony is argued in Chapter 6. Chapter 7 argues that communities can be loci of epistemic authority. Chapter 8 defends the possibility of moral authority—epistemic authority with respect to beliefs in the moral domain. Chapter 9 defends the possibility of authority over religious belief. Chapter 10 discusses the interaction of trust and disagreement. Chapter 11 seeks to connect epistemic authority with autonomy and to defend the latter against debunking arguments.

The book contains much that’s worthy of critical attention. However, I will confine my discussion here to what I take to be Zagzebski’s core argument, for the rationality of trust in oneself and the consequent rationality of trust in others. (Chapters 2–3)

2. *Extreme epistemic egoism*

Let me begin with an overview of the argument. The target of the argument is a position Zagzebski calls *extreme epistemic egoism*: “The most extreme form of epistemic self-reliance is one in which someone refuses to take the fact that someone has a given belief as a reason to believe it herself.” (p.52) Zagzebski aims to show that one cannot rationally adopt the position. The argument proceeds in three stages. In the first, Zagzebski provides a working account of rationality. In the second, she argues that, given the working account of rationality, trust in the reliability of one’s epistemic faculties is rational. In the third, she argues that, given that trust in the reliability of one’s own faculties is rational, trust in the reliability of others’ faculties is rational. I shall consider the three stages in order. In assessing the argument, we should keep in view her governing aim, to exploit grounds that “almost all modern philosophers would accept.” (p.2)

2.1. *Rationality*

Zagzebski’s guiding account of rationality is the following. “To be rational is to do a better job of what we do in any case—what our faculties do naturally.” (p.30) There are various things that we, and our faculties, do naturally. Being rational is a matter of doing those things better. According to Zagzebski, “This means that there is a connection between the natural and the normative, in particular, a connection between the self as it naturally operates and the way it should operate.” (p.33) Presumably, Zagzebski doesn’t hold that most modern philosophers accept that as a substantive claim about rationality. Since she doesn’t defend it, we should treat it as a stipulative definition. So understood, her aim will be to defend the claim that doing what we naturally do better involves trusting others.

In order to determine the dictates of rationality, as Zagzebski understands them, one would need to ascertain three things. First, one would need to establish what it is for something to be natural as opposed to non-natural. As far as I can tell, Zagzebski doesn’t explain the distinction, although she seems to associate the natural and the prereflective. Second, with the distinction in hand, one would need to establish which of the things we do are natural and which are not. Zagzebski makes various claims about things we do, or might do, to the effect that they are, or are not, natural. But as far as I can tell, she doesn’t defend those claims. It would be natural to view claims about naturalness as partly empirical claims, dependent on work in the natural sciences. However, Zagzebski doesn’t present empirical evidence in support of her claims, so it may be that she understands the natural–non-natural distinction in a way that is not hostage to work in the natural sciences. Third, one would need to provide an account of the natural things that we do that was able to sustain claims about what it would be to do those things better. One way of proceeding here might be to defend (i) claims about the (natural) functions of the things that we naturally do and (ii) claims

about how we might proceed in order to perform those functions better. Again, as far as I can tell, Zagzebski doesn't attempt to do this.

The central figure in Zagzebski's account of rational psychology is *dissonance*. Beliefs, emotions, desires, and decisions can conflict with one another. We experience conflict amongst our mental states as dissonance. It seems to me that the experience of dissonance is basic. It cannot be explained or analysed in terms of some other experience. (p.29)

Sometimes our mental states conflict, as for example when one believes p and believes *not-p*. The experience of dissonance affords us our most basic mode of access to the obtaining of conflict. Potential to induce dissonance may even be what makes it the case that mental states conflict. According to Zagzebski, "We naturally desire and attempt to achieve a harmonious self." (p.31) That is, we naturally desire and attempt to achieve a self that is free from (the experience of) dissonance. The experience of dissonance, together with the desire not to experience it, compels us to seek resolution, typically by shedding at least one of the conflicting states. Dissonance figures centrally in Zagzebski's discussion.

Zagzebski claims that the desire to be free from dissonance is natural. Presumably, she also holds that the experience of dissonance, or the capacity appropriately to suffer the experience of dissonance, is natural. As noted above, such claims about naturalness are hard to assess, especially in the absence of supporting evidence.

Setting issues about naturalness to one side, the basic functions of dissonance and the desire that it be resolved raise further questions. One set of questions concern the reliability of the capacity to experience dissonance. Is the capacity taken to be such that, whenever there is conflict amongst one's mental states, one will experience that conflict as dissonance? Zagzebski suggests that conflict is not in that sense self-intimating. There can be conflicts amongst our beliefs and desires without our thereby experiencing dissonance. (p.31) More importantly, is the capacity infallible, so that one experiences dissonance only in cases in which there really is conflict amongst one's attitudes? Zagzebski appears not to address the question. Whether or not the capacity is fallible, further questions attend the range of admissible responses available to us in cases in which we experience dissonance.

Zagzebski seems to hold that in cases in which a pattern of mental states gives rise to an experience of dissonance, our cognitive ends will best be served by shedding at least one of the apparently conflicting mental states. But if dissonance is fallible, there may be some cases in which we would be better off living with the experience, as we do with some other experiences that we take to be illusory. A further question at this point is whether the experience of dissonance is intrinsically disvalued, perhaps being akin to a form of cognitive discomfort. As we've seen, Zagzebski appeals to a desire for cognitive harmony in addition to the basic experience of dissonance. That she does so suggests that she thinks that the desire is required in order for the experience of dissonance to motivate resolution. And that suggests that, on her view, the experience of dissonance per se is not intrinsically disvalued. In that case, we might wonder whether it is the general desire for harmony that should be shed, perhaps in favour of a more nuanced desire to resolve dissonance only in those cases in which it is genuinely revelatory of conflict. Zagzebski claims that the desire for harmony is natural. Perhaps, then,

she holds that retaining it is a component of being rational. However, if being rational is doing *better* at what we do naturally, it might be argued that it would be better to move from the less to the more nuanced desire. Finally, suppose that the experience of dissonance is intrinsically disvalued. In that case, shedding the desire for harmony wouldn't suffice for harmony. However, it would be natural to wonder whether it's possible to shed individual experiences of dissonance themselves, especially if we believe that such experiences may, on occasion, be misleading. (One might consider here our initial and more reflective responses to certain putative paradoxes. The initial experience of dissonance appears sometimes to give way in the face of reflection in a way that may, but doesn't obviously, depend upon losing or gaining further mental states other than the feeling itself.)

Whatever the basis for so judging, we sometimes judge that we ought to resolve dissonance in one or another way. Should we act on our judgments? Zagzebski writes:

In cases in which we do not do what we judge we ought to do to resolve dissonance within the self, we are not rational because it is not rational to fail to do what we judge we ought to do. (p.32)

It's not clear here whether Zagzebski's claim is driven in turn by her views about what it would be to do better what we do naturally. However, her claim seems questionable. Our judgments about what we ought to do are presumably made correct or incorrect by what we in fact ought to do. In that case, suppose that I ought to resolve dissonance in way *X* rather than way *Y*, but judge—incorrectly—that I ought to resolve dissonance in way *Y* rather than way *X*. Resolving dissonance in accord with my judgement seems *less* rational than failing to do so. However, since I've judged that I ought to resolve dissonance in way *Y* rather than way *X*, it would induce dissonance, and to that extent count as less rational, to resolve dissonance in way *X* rather than way *Y*. But whether acting in accord with my judgment counts as less rational than the alternatives *overall* depends on whether the dissonance between the judgment and what I do is worse than the dissonance induced by incorrectly seeking to resolve the initial dissonance in way *Y* rather than way *X*. Moreover, and more generally, the potential dissonance induced by acting out of accord with my judgment could have been avoided by dropping the judgment. At best, then, what we have are reasons to think that it would be worse to (judge that we ought to *A* and fail to *A*) than to (shed the judgment that we ought to *A* or act in accord with that judgment). But in cases in which the judgement is erroneous, it might be better to shed the judgment than to act on it.

Let me summarise the discussion to this point. It would be useful to have a more developed account of what it is for what we do to be natural, and of what it would be for us to do it better. It would also be useful to have a more developed account of the psychological machinery to which Zagzebski appeals. The next section considers the question whether it would be rational to trust our epistemic faculties—that is, whether doing so would be a matter of doing better what we naturally do.

2.2. *Self-trust*

Zagzebski's aim is to argue that self-trust—trust in the general reliability of our faculties—is reflectively rational. Her argument has two stages. In the first stage, she argues that self-trust is natural or prereflective. In the second, she argues that amongst the reflective, there is no superior alternative to self-trust.

Zagzebski begins the first stage with an assumption.

...I assume that with possibly a few exceptions, everyone desires truth, and it is natural for everyone, even if there are a few who do not desire it. (p.34)

Zagzebski doesn't defend the claim. Furthermore, as noted above, she admits that contemporary philosophers "...generally do not speak of natural desires..." (pp.33–34) It would be good to have been presented with reasons to think that the claim is, or should be, accepted by contemporary philosophers. Furthermore, it's not entirely clear what the claim amounts to, or what would satisfy the purported desire. Is it a desire for all truths, or only some? And if the desire is for only some truths, is there a general characterisation of the distinction between those we desire and those we don't? What would satisfy the desire for a given truth, say the answer to a specific question? Would it suffice if I were to come to believe the correct answer to the question? Or must I in addition acquire justification or knowledge? Zagzebski doesn't address such questions, and I was unable to reconstruct adequate answers.

Zagzebski continues:

In addition to the natural desire for truth, I assume that there is a natural belief that the natural desire for truth is satisfiable. (p.35)

Again, the claim is not further articulated or defended. Again, we might wonder about the content of the purported belief. Does it amount to the belief that, with respect to any question whose answer I desire, I can come to *know* the answer to that question? Whether or not it is natural, such a belief seems immodest. We might think it would be better not to harbour such a belief. Alternatively, perhaps the belief is weaker, with consequent effects on the remainder of Zagzebski's argument. Whatever the precise content of the assumed belief, it figures as the basis of the following argument:

...if I think I can get truth, I must think I can sometimes get it without the aid of reflection, and probably do so often. But if so, I must have a basic trust in the connection between my faculties and the truth. Furthermore, since I relied upon my faculties to answer my questions before I started reflecting, trusting my faculties was already part of the pre-reflective self. (p.36)

Given the assumption that I naturally believe that I can get the truth, it must be that I naturally, or prereflectively, trust my faculties to furnish me with truths—that is, with true beliefs, or with beliefs meeting some other standard. Obviously, the strength of the conclusion depends on the strength of the assumption that supports it. If the assumption were that I believe that I *sometimes* form beliefs that are true, the conclusion would be that I trust my faculties to not lead me astray in *all* cases. On stronger assumptions—assumptions that would then impose greater burdens on their defender—we might reach stronger conclusions.

Zagzebski's draws on her intermediate conclusion about my trust in my faculties to derive further conclusions with potential relevance to questions about trust in others' faculties:

The faculties I rely on in forming beliefs operate on an environment, so trusting my faculties includes trusting that the environment is appropriate to the faculties. My faculties may operate on the environment directly, or they may operate indirectly through the faculties of others. It is natural to believe what people tell me. Trust in my faculties and environment includes trust in the faculties of many other persons. Trust in others, like trust in the self, is the starting point. It is a component of the prereflective self. (p.38)

Again, Zagzebski's conclusion depends on an undefended claim about what is natural—in this case, the claim that it is natural to believe what people tell me. Furthermore, the claim is underspecified. Is the claim that it is natural always to believe what we are told, or only sometimes? If the claim is that we only sometimes believe what we are told, can the distinction between cases in which we do believe what we are told, and cases in which we don't, be further specified?

Zagzebski's argument to this point is difficult to assess. She makes a number of claims, or assumptions, about what is natural without attempting either to specify their precise content or to defend them. Let's suppose, however, that those claims are accepted. It is natural to trust our faculties and, moreover, to trust other peoples faculties. Have we defended the rationality of either form of trust? Not yet. Being rational is not doing only what we do naturally. Rather, it's doing better what we do naturally. Reasons are required for holding that doing better what we do when we naturally trust will involve trusting either our own faculties or others'. We require reasons to accept that doing better what we naturally do will involve preserving, perhaps in improved form, our natural trust in ourselves and others, rather than replacing it with one or another form of scepticism.

Zagzebski treats the question whether self-trust is a component of our doing better as the question whether self-trust can, or should, survive reflection. This procedure appears to rely on the assumption that reflection is reliable. One alternative would be to allow that we might not be in a position to assess what it would be for us to do better what we naturally do and that our best attempts to do better might result in our doing worse. However, setting aside qualms about the procedure, her answer to the question is that self-trust should survive reflection. Her defence of that answer draws on recent discussions by Richard Foley (1991) and William Alston (2005) of traditional foundational projects. As she puts the upshot of those discussions,

A number of philosophers [including Foley and Alston] have observed that there is no noncircular way to determine that the natural desire for truth is satisfiable, or to put the claim in the preferred idiom, there is no noncircular way to tell that our belief-forming faculties are reliable as a whole. (pp.38–39)

On the basis of the purported observation—for short, the observation of epistemic circularity—Zagzebski argues as follows:

Is it rational to have self-trust after reflection? ... Reflective self-trust resolves the dissonance we have when we discover epistemic circularity, and that seems to me to be rational. It is rational to believe that my faculties are trustworthy for the purpose of getting the truth; it is rational to treat my faculties as if they will get me to the truth... Of course, if someone thinks it is not rational to have a belief without noncircular justification, then self-trust is not rational, but then none of our beliefs is rational. However, I see no reason to think that that is what rationality is. (p.43)

This passage presents the core of Zagzebski's argument for the reflective rationality of self-trust. The alleged discovery that there is no noncircular way to tell that our belief-forming faculties are reliable as a whole leads to an experience of dissonance. The experience of dissonance is indicative of conflict amongst our mental states. If we assume that the experience is in this case veridical, it reveals that the belief in epistemic circularity conflicts with another state. Plausibly, the state with which it conflicts is our desire for global noncircular justification. Reflection either generates, or makes salient, that desire. The optimal resolution of the conflict would involve shedding the desire for global noncircular justification. (According to Zagzebski the alternative of adopting one or another form of scepticism would not amount to an improvement on what we anyway do.) Self-trust is either the outcome of shedding the desire, or is what enables us to shed the desire. The upshot of reflection is that doing better what we naturally do leads to the adoption, or preservation, of self-trust.

I have three main concerns with this argument. The first concern is that depending in part on what precisely it would take to provide global noncircular justification, it is not obvious that the arguments to its impossibility presented by Foley and Alston are successful. The second concern is that Zagzebski's argument depends on the assumption that the alternatives Zagzebski considers—self-trust, in Zagzebski's proprietary sense, and global noncircular justification, in whatever sense this is supposed to be demonstrably impossible—are exhaustive. However, it is far from obvious that the assumption is correct. For example, it isn't obvious that a global noncircular justification of the general reliability of our faculties is required in advance of, and so cannot depend upon, the individually warranted issue of those faculties. Whether that is so, and whether there are other reasonable alternatives, can only be decided by more detailed reflection on the alleged motivations for noncircular justification and the range of possible responses to them. (For example, addressing the question whether or not such a justification would be impermissibly circular would depend on further specification, and defence, of the ban on circularity.)

The third concern arises from the question whether the desire for global noncircular justification is grounded in what we naturally believe and desire. Suppose it is. In that case, we would have an explanation for why people have the desire and for why some of them would rather endorse scepticism than give it up. But in that case, it becomes pressing to explain how shedding such a natural desire could be a component of doing better what we do naturally. Furthermore, the desire appears to conflict with self-trust, suggesting either that self-trust isn't natural, or that we are by nature conflicted. Alternatively, suppose that the desire isn't natural. In that case, we would need a different explanation for the ease and tenacity with which the desire is acquired. Moreover, given the naturalness of self-trust, we would need to explain why it hasn't been more widely perceived that the desire can easily be shed in its favour.

In this section, I've indicated some apparent weak points in Zagzebski's argument that self-trust is a component of doing better what we do naturally. Zagzebski's claims about what is natural are difficult to assess in the absence of further specification and defence. Her argument from the alleged impossibility of global noncircular justification to the rationality of self-trust is underdeveloped. In

the final section, I'll consider Zagzebski's argument for the rationality of trust in others, built on foundations provided by her defence of self-trust.

2.3. Trust in others.

Zagzebski's basic statement of the argument that trust in others is rational is brief:

As a self-reflective person, I trust the beliefs I form when I am epistemically conscientious, which I defined as using my faculties in the best way I can in order to satisfy the desire for truth. 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 general trust in myself and I accept the principle that I should treat like cases alike, I am rationally committed to having general trust in them also. (p.55)

My trust in the general reliability of my own faculties isn't grounded in independent (noncircular) evidence of their reliability, but rather in the fact that, given my natural desire for truth, it is better for me to believe that my faculties are reliable than not to. I can't provide a global noncircular justification of the reliability of my faculties. Hence, the only alternative to self-trust would be scepticism, which would conflict with my natural desire for truth. Self-trust requires that I believe that my faculties are reliable. But I have standing reason to treat like cases in the same way. Hence, if I have reason to hold that others are relevantly like me, then I should hold that other people's faculties are reliable too.

As it stands, this argument seems unconvincing. My belief that my faculties are reliable is grounded not in evidence of their reliability, but rather in the fact that believing so will better serve my desire for truth. But that desire isn't served by my believing in addition that anyone relevantly like me will also have reliable faculties. The desire for truth doesn't give me any reason to believe that others like me—even others with faculties like mine—also have reliable faculties. The desire doesn't itself require me to believe, in addition to believing that I am reliable, that my reliability is explained by a shareable property of mine. It requires only that I believe my faculties to be reliable. Furthermore, it's presumably a good thing that it doesn't impose such additional demands, since I have no evidence that they are met.

Zagzebski seeks to clarify her argument in the following passage:

...the crucial point is that if, in believing in a way I trust, I come to believe that other persons have the property in virtue of which I trust myself, then I am committed to trusting them also. (p.57)

On one reading of the passage, the argument would be this. I trust myself in virtue of my possession of a particular property. According to the argument for self-trust, the operative property is my desire for truth. Consequently, if I discover that others have the same property—that is, the property of desiring truth—then I am rationally committed to trusting them. However, when the passage is read in that way, the argument fails. At best, discovering that another desires truth might commit me to holding that they, like me, should trust their own faculties. It gives me no reason to trust their faculties, since trust in their faculties doesn't figure in satisfying my desire. I assume, therefore, that that isn't the intended reading.

An alternative reading would be this. I am rationally committed to self-trust. However, I can't simply trust myself. Rather, I must make my trust intelligible given my other beliefs. My other beliefs include the belief that if I am reliable, then my reliability must be grounded in a shareable property of mine. That shareable property would be "the property in virtue of which I trust myself". (p.57) Thus, my self-trust requires that I believe that I have the shareable property in virtue of which I am reliable—a property my possession of which suffices for reliability. Now suppose that in acquiring beliefs in a way that I trust, I come to believe that another has the same property. Since I'm committed to treating like cases in the same way—so that, for example, I'm unwilling to accept that the property suffices for reliability only when it's possessed by me—I'm committed to taking the other as reliable in virtue of their possession of the property.

Read in this way, Zagzebski's argument faces a dilemma. On the first horn, suppose that my only mode of access to the alleged shareable property is via its being the property in virtue of which I'm reliable. In that case, we can perhaps allow that it's reasonable for me to believe that I have the property. The difficulty is in then explaining how I can come to believe that others have the property, at least in advance of coming to believe that they are reliable. What else is supposed to ground my belief that others are like me in this respect?

Alternatively, on the second horn, suppose that I form a specific conception of a property of mine and hypothesise that it's my possession of that property that explains my reliability. Now I might come to have reason to believe that others have the specific property. But in that case, the question whether they are reliable or not will figure in assessment of the hypothesis that possession of the property suffices for reliability. And it seems that I have no special reason to believe that the hypothesis is correct—that because they have the property, they are reliable. Rather, it would seem equally reasonable for me to believe that they have the property and are not reliable, and so to withhold assent from the hypothesis that the property is responsible for my reliability.

In this section, I've indicated some apparent weaknesses in Zagzebski's argument that it's rational to trust in the reliability of others' faculties. The main issue here is the difficulty in seeing how the reasons for self-trust to which Zagzebski appeals can be translated into reasons for trusting others. There is much more to say both for and against Zagzebski's argument. And there is yet more to say about the remainder of Zagzebski's rich and rewarding book.

References

- Alston, W. 2005. *Beyond Justification: Dimensions of Epistemic Evaluation*. Ithaca, NY: Cornell University Press.
- Foley, R. 2001. *Intellectual Trust in Oneself and Others*. Cambridge: Cambridge University Press.
- Raz, J. 1988. *The Morality of Freedom*. Oxford: Oxford University Press.