Resistance to Evidence
Triggers and Epistemic Status

This chapter dwells at the intersection of the social psychology of knowledge resistance and epistemic normativity to offer the first full taxonomy of resistance to evidence. It first individuates the phenomenon via paradigmatic instances, and then it taxonomises it according to two parameters: (1) paradigmatic triggering conditions and (2) epistemic normative status. I argue that the phenomenon of resistance to evidence is epistemologically narrower but psychologically broader than is assumed in the extant literature in social psychology. This, in turn, gives us reason to believe that addressing this phenomenon in policy and practice will be a much more complex endeavour than is currently assumed. In the remainder of the book, I examine the extant literature on evidence, defeat, justification, permissible suspension, and epistemic responsibility in search of the normative resources required to fully accommodate the psychological breadth and epistemic normative status of the phenomenon of resistance to evidence.

1.1 Resistance to Evidence

The notion of resistance to evidence, while subject to thorough investigation in social psychology,[1] is acutely under-theorised in the philosophical literature. As a result, we are still to understand the normativity of the resistance phenomenon: What is (epistemically) wrong with resistance to evidence? What are its triggers? How does the normativity of resistance to evidence interact with norms of inquiry and the epistemic justification of belief?

Consider the following cases:

Case #1. **Testimonial Injustice:** Anna is an extremely reliable testifier and an expert in the geography of Glasgow. She tells George

---

[1] See, e.g., Kahan 2013, Klintman 2019. See also the 'Knowledge Resistance' multidisciplinary research project at Stockholm University: https://knowledge-resistance.com.

9

that Glasgow Central is to the right. George believes women are not to be trusted; therefore, he fails to form the corresponding belief.

Case #2. **Political Negligence:** Bill is a stubborn supporter of President Dump. In spite of all evidence that is readily available to him (via mainstream media, Dump's own actions and public statements, etc.) suggesting that Dump is a bad president, Bill stubbornly refuses to believe that Dump is a bad president.

Case #3. **Science Scepticism:** Neda is an anxious cogniser; in particular, she is very careful when it comes to accepting science communication: whenever well-recognised, reliable experts assert that anthropogenic climate change is occurring or that vaccines are safe, Neda suspends belief thinking, 'Well, scientists sometimes get it wrong! I'll do my own research.'

Case #4. **Perceptual Non-responsiveness:** Alice is looking straight at the table in front of her and fails to form the belief that there's a table in front of her.

Case #5. **Unwarranted Optimism:** Mary is an optimist. When her partner Dan spends more and more evening hours at the office, she's happy that his career is going so well. When he comes home smelling like floral perfume, she thinks to herself: 'Wow, excellent taste in fragrance!' Finally, when she repeatedly sees him having coffee in town with his colleague Alice, she is glad he's making new friends.

Case #6. **Misdirected Attention:** Professor Racist is teaching college-level maths. He believes people of colour are less intelligent than white people. As a result, whenever he asks a question, his attention automatically goes to the white students, such that he doesn't even notice the Black students who raise their hands. As a result, he believes Black students are not very active in class.

Case #7. **Friendly Detective:** Detective Dave is investigating a crime scene. Dave is extremely thorough but, at the same time, a close friend of the butler. Dave finds conclusive evidence that the butler did it – the butler's gloves covered in blood, his fingerprints on the murder weapon, a letter written by the butler confessing to the crime – but he fails to form the corresponding belief: Dave just can't get himself to believe that his friend would do such a thing.

What is going on in these cases? Note that they involve very different sources of knowledge (e.g. testimony, perception, inductive inference) and that the failures at stake come about for very different reasons (e.g. prejudice, motivated reasoning, epistemic anxiety, lack of attention, partisanship, bias, wishful thinking). All of these are bad things, epistemically, in their own right. At the same time, the cases also have one important feature in common: for all these subjects, there is excellent evidence easily available to them, which they fail to take up.

Several philosophers have offered source-bound diagnoses of particular incarnations of this phenomenon (in terms of, e.g., epistemic injustice (Fricker 2007), disregard for the nature and/or normativity of telling (Moran 2006, Hazlett 2017), breach of norms of attention (Siegel 2017)), but very few have tried to offer an overarching explanation of what they all have in common. However, once we look at these cases together, it becomes clear that, on top of the case-specific problems, they plausibly exhibit a common variety of epistemic failure: resistance to easily available evidence.

1.2   The Social Psychology of Evidence Resistance

1.2.1   *Evidence Resistance and Motivated Reasoning*

A predominant hypothesis in social psychology (e.g. Lord et al. 1979, Taber and Lodge 2006, Molden and Higgins 2012, Kahan 2013, Kahan et al. 2016) that seeks to explain 'knowledge resistance' (i.e. resistance to acquiring easily available knowledge) principally does so with reference to politically motivated reasoning. Under the banner of this wider hypothesis, we find various research results that have been taken, in various ways, to support the view that a thinker's prior political convictions (including politically directed desires and attitudes about political group membership) best explain why they are inclined to reject expert consensus when they do (Kahan et al. 2011, Kahan 2013).

Early studies in the psychological literature that set the groundwork for this explanatory thesis focused initially on how political ideology influences the evaluation of evidence. For example, Lord et al. (1979) report a study in which subjects were provided with the same set of arguments for and against capital punishment and were asked to assess the strength of these arguments. Subjects' assessment of the strength of the arguments then strongly correlated with their existing views about the rights and

wrongs of capital punishment. In short, subjects *already disposed* to object to capital punishment were more persuaded by the arguments against it, and the opposite was the case for those initially predisposed to favour capital punishment. (See also Kunda 1987 for discussion of how political ideology seems to have a bearing on causal inference patterns.)

A second wave of research in this area, led largely by Dan Kahan and his colleagues, has suggested that political ideology not only influences how we think about the persuasiveness of arguments for and against those ideologies themselves but also that our inclination to accept (or reject) scientific consensus across a range of areas is highly sensitive to what political ideology we already accept. For example, Kahan and his collaborators present studies aimed at demonstrating that background political ideology impacts whether we align with or go against expert consensus on topics ranging from global warming to the safety of nuclear power (Kahan et al. 2011, Kahan 2014, Kahan et al. 2016; cf. Carter and McKenna 2020).

In light of this second wave of research, the received thinking about resistance to evidence takes such resistance to be principally a manifestation of politically motivated reasoning (Kahan 2013).

This position, while widely discussed in social psychology, has received comparably less attention in philosophy. Furthermore, typically, philosophers who have discussed it have explored the consequences of this empirical hypothesis while taking its merits at face value (e.g. Ancell 2019, Carter and McKenna 2020).

However, on closer and recent inspection, the hypothesis is both empirically and epistemically problematic. Empirically, there are worries that, in extant studies, political group identity is often confounded with prior beliefs about the issue in question; and, crucially, reasoning can be affected by such beliefs in the absence of any political group motivation. This renders much existing evidence for the hypothesis ambiguous (Tappin et al. 2021).

Epistemologically, the worry is that the hypothesis is ineffective for making crucial distinctions among a number of phenomena, such as (1) concerning epistemic status: between epistemically impermissible resistance to evidence, on the one hand, and justified evidence rejection, on the other – after all, if the extant priors that are correlated with political group identity are justified priors and if evidence resistance is sourced in these justified priors rather than in motivated reasoning, we will have failed to distinguish justified evidence rejection from unjustified evidence resistance; and (2) concerning triggers: between instances of motivated

reasoning, on the one hand, and epistemically deficient reasoning featuring cognitive ('cold') biases and unjustified premise beliefs, on the other.

Furthermore, difficulties in answering the question as to what triggers resistance to evidence have very significant negative impacts on our prospects of identifying the best ways to address this phenomenon and to avoid its unfortunate practical consequences. If resistance to evidence has one main source – for instance, a particular type of mistake in reasoning, such as motivated reasoning – the strategy to address this problem will be unidirectional and targeted mostly at the individual level. In contrast, should we discover that a pluralistic picture is more plausible when it comes to what triggers resistance to evidence – whereby this phenomenon is, for example, the result of a complex interaction between social, emotive, and cognitive phenomena – we would have to develop much more complex interventions at both individual and societal levels.

1.2.2   *Evidence Resistance and Epistemic Vigilance*

One noteworthy way in which knowledge resistance manifests is in the context of a hearer's receipt of testimony from a speaker. Two kinds of examples which have received particular attention include cases of (1) resistance to expert testimony (e.g. widespread resistance to scientific evidence about climate change, as well as during the onset of the COVID-19 pandemic; Kearney et al. 2020), and (2) resistance to testimony from marginalised groups, which provides the central point of reference in the literature on testimonial injustice (Fricker 2007). In both kinds of cases, the hearer's response to testimony is epistemically defective.

An important strand in the social psychology of testimonial knowledge transmission suggests that the above phenomena could be explained via the misfiring of an otherwise beneficial epistemic vigilance mechanism. Research by Dan Sperber and colleagues (2010) and related work by Hugo Mercier (2020) suggest that the risks that we as testimonial recipients face in being accidentally or intentionally misinformed are ones that we are well positioned to navigate via a suite of cognitive mechanisms of epistemic vigilance for sorting, sifting, and discerning information coming from other human beings (whether immediately or mediately). It is this suite of mechanisms that is postulated, on the epistemic vigilance programme, as important in explaining both the honesty of speakers and the reliability of their testimony.

If Sperber et al. (2010) and Mercier (2020) are right and we do benefit from a suite of mechanisms that make us epistemically vigilant, the

phenomenon of resistance to evidence may be explained as an instance of misfiring of our epistemic vigilance mechanisms. If these vigilance mechanisms are misfiring, they will lead us to respond with distrust and disbelief when trust and belief are the appropriate responses. In this way, epistemic vigilance may lead to resistance to evidence. One explanation for this might lie with the fact that we now inhabit a very different epistemic environment from the environment that our mechanisms for epistemic vigilance evolved in: recent technological advances have placed us in the midst of information (and misinformation) overload. Since our cognitive mechanisms of vigilance, the thought would go, have not evolved in such a heavyweight informational environment, they are misfiring in an attempt to cope.

Yet, a wave of research on deception recognition paints a mostly pessimistic picture about the plausibility of the very existence of vigilance mechanisms in us. A wide range of studies testing our capacities for deception recognition show that we are very bad at it: our prospects of getting it right barely surpass chance (e.g. Kraut 1980, Vrij 2000, Bond and DePaulo 2006). To see just how well established this result is in the relevant psychological literature, consider the following telling passage from Levine et al. (1999, 126): 'the belief that deception detection accuracy rates are only slightly better than fifty-fifty is among the most well documented and commonly held conclusions in deception research'.

Crucially, it is not hard to see that if these studies are right and we detect deception with an accuracy rate that is barely above chance, both the hypothesis that we have evolved cognitive mechanisms for epistemic vigilance to help us secure the reliability of testimonial exchanges and the idea that resistance to evidence is the result of our vigilance mechanisms misfiring become rather implausible.

More recently, though, some voices in the deception detection literature have grown disenchanted with the received view on the issue. In particular, J. Pete Blair et al. (2010) argue that the past forty years of research in deception detection have neglected the role of contextual clues. According to them, accuracies significantly higher than chance can be consistently achieved when hearers are given access to meaningful contextual information. On the face of it, this seems like it might be the sort of result vigilance champions need to establish that the vigilance mechanisms make the needed difference for testimonial entitlement (i.e. by increasing reliability). The vigilance mechanisms, the thought would go, have evolved to work in conjunction with the contextual information Blair et al. discuss.

Unfortunately, though, upon closer examination, these results will not do the trick for the epistemic vigilance champion. To see why, it is

important to look more closely at the type of contextual information that has been given to the subjects for the purposes of this study and ask the question: 'How plausible is it that this kind of information (i.e. information that is shown to increase reliability in deception detection) is the kind of information that, when had, would still require us to have extra input from our vigilance mechanisms given the context?' After all, if the study gives information such as 'This is a reliable testifier', this is the kind of information that seems to justify testimonial belief on its own – it's simply evidence that the testifier is telling the truth. Conversely, if the study provides the subject with evidence that the testifier in question is unreliable, again, one need not host epistemic vigilance mechanisms in order to justifiably withhold belief.

The Blair et al. study identifies three types of what they dub 'contextual content' that raise the success rates for deception detection (2010, 424–425): (1) contradictory content – for example, a testifier claims to have been at home on a given night, but the hearer is told by a trusted source that she saw the testifier out at a restaurant on the night in question; it is likely that the testifier's statements will be flagged as deceptive; (2) statistically normal content – for example, knowledge about the testifier's normal activities; if the testifier's statements or performance are implausible given this statistically normal information, the statements are more likely to be flagged as potentially deceptive; and (3) information that increases the perceived probability of deceit – for example, a situation in which a number of shortages have occurred at a bank, but the shortages stop when one of the employees goes on vacation and begin again when the employee returns; this information may cause an interviewer to believe that the employee's statements are deceptive.

These results are, of course, hardly surprising, either empirically or epistemologically: it seems trivially true that, if given the right kind and amount of contextual information in advance, most of us should be and are able to go so far as to be impeccable deception detectors on mere garden-variety epistemic grounds – no extra mechanisms needed. As a limit case, if I know in advance that everybody is lying, for instance, I will likely be very good – indeed, infallible – at detecting deceit. What matters for us here, however, is whether the kind of information that does the trick in the study at hand is the kind of information that would plausibly increase the general reliability of our vigilance mechanisms rather than deliver sufficient evidence for/against a particular piece of testimony on its own. The plausible answer, however, I contend, is clearly the latter: no special vigilance-like psychological skills are required in these cases, as the

evidence is enough to justify the response. Furthermore, and interestingly, one out of three Blair et al. experiments failed to confirm their hypothesis (2010, 427): this was the experiment that gave participants the most limited and subtle contextual information. Thus, the experiment that most closely resembled a garden-variety testimonial exchange, where the hearer does not have a whole lot of antecedent knowledge about the speaker, failed to deliver high rates of successful deceit detection. This, again, does no look very promising for the vigilance hypothesis.

If this is right – if the hypothesis that we host special epistemic vigilance mechanisms is implausible to begin with – then the hypothesis that instances of resistance to evidence are instances of our vigilance mechanisms misfiring remains unvindicated as well.

1.3   Rejecting Evidence: A Taxonomy

What we have seen so far is that the extant research on evidence resistance suffers from both empirical and epistemological shortcomings in identifying the triggers behind the target phenomenon: on the one hand, epistemologically, we need to distinguish between unjustified evidence resistance – sourced in all kinds of epistemically impermissible belief/suspension formation, such as motivated reasonings, biases, etc. – and epistemically justified evidence rejection – sourced in justified prior beliefs. On the other hand, even when zooming in on epistemically problematic instances of the phenomenon it is not clear how much evidence resistance is sourced in cold rather than hot biases or in updating on unjustified priors rather than biases.

These difficulties in answering the question as to what triggers resistance to evidence have, in turn, very significant negative impacts on our prospects of identifying the best ways to address resistance to evidence. If resistance to evidence has one main source – for instance, a particular type of mistake in reasoning, such as motivated reasoning – the strategy to address this problem will be targeted at the individual level. In contrast, should we discover that a pluralistic picture is more plausible when it comes to what triggers resistance to evidence, we would have to develop much more complex interventions at both individual and societal levels. Finally, if it turns out that the vast majority of instances of alleged evidence resistance are actually explained by epistemically justified evidence rejection – say, because cognisers find themselves in environments polluted with misleading defeaters for the evidence at stake – our interventions should only target the relevant epistemic environment rather than any particular cogniser or belief-formation mechanisms.

Of course, the question as to what is actually happening, on the ground, in evidence-resistant communities and individuals is to be answered by careful, epistemologically informed empirical studies. It is not my ambition to settle this question from the armchair – nor should any epistemologist attempt to do so.

The ambition of this book is to offer the first epistemology of evidence resistance that can inform future empirical studies on the topic. To this effect, I will start by putting forth in Table 1.1 a simple taxonomy of the phenomenon in order to isolate the epistemically problematic instances that we are interested in.

In theory and practice, it is crucial, before undergoing an epistemological analysis of problematic cases of evidence resistance and before testing an empirical hypothesis having to do with the instantiation of this problematic phenomenon in a particular community, to first distinguish evidence resistance from its epistemically benign cousin: justified evidence rejection. One reason why this is crucial has to do with addressing the phenomenon in policy and practice: depending on whether we are dealing with justified evidence rejection or epistemically impermissible evidence resistance, different interventions are warranted. For combatting epistemically justified evidence rejection, engineering enhanced social epistemic environments should do the trick: since we are dealing with reliably epistemically responsive agents, we can rely on them to update in line with a non-polluted epistemic environment. This will likely require combatting rebutting defeaters via evidence flooding: evidence-resistant communities, inhabiting polluted epistemic environments, cannot be reached via the average communication strategies designed to reach the mainstream population inhabiting a friendly epistemic environment (with little to no misleading evidence). What is required is quantitatively enhanced reliable evidence flow (more evidence in favour of the scientifically well-supported facts will, in rational agents, work to outweigh the misleading evidence they have against the facts), as well as qualitatively enhanced reliable evidence flow (evidence from sources that the agent trusts – that are trustworthy vis-à-vis the agent's environment). Furthermore, to combat mistrust in reliable sources, quantitatively and qualitatively enhanced evidence aimed at combatting undercutting defeat (misleading evidence against the trustworthiness of reliable sources) will be needed. One straightforward way to do this is by flooding evidence-resistant communities with evidence from sources they trust in favour of the trustworthiness of sources they fail to trust due to misleading undercutting defeaters.

Table 1.1 *Evidence rejection: a taxonomy*

Evidence rejection							
Epistemically justified				Epistemically unjustified (*evidence resistance*)			
Via rebutting epistemic defeat		Via undercutting epistemic defeat		Via unjustified doxastic defeat		Independent of doxastic defeat	
Not misleading	Misleading	Not misleading	Misleading	Via proper updating	Via improper updating	One-off/ isolated	Dispositional
							Sourced in cold bias / Sourced in hot bias

18

In contrast, for combatting cases of unjustified evidence resistance agent-based interventions will be needed: for example, increasing the availability of cognitive flexibility training (e.g. in workplaces and schools, alongside anti-bias training) will be among the more efficient interventions. Cognitive flexibility training helps with enhancing open-mindedness to evidence that runs against one's held beliefs and to alternative decision pathways (Garner 2009, Griffin et al. 2012).

1.3.1   *Justified Evidence Rejection*

Let's take the science sceptic case as our toy case to illustrate. In the original variation of the case, of course, Neda was evidence resistant *tout court* due to her epistemic anxiety. The point I am trying to impress on you, however, with the above taxonomy, is that not all science sceptics need be like Neda: they need not be unjustifiably nor irrationally rejecting scientific evidence. A science sceptic Neda* could be rejecting scientific testimony about, for example, the safety of vaccines because her environment is polluted with misleading defeaters: say, she lives in a community where an overwhelming majority of testimony that she gets suggests that vaccines are not safe. Say, also, that these testifiers are otherwise reliable testifiers, with an impeccable track record (who just happen to get things wrong on this particular occasion – after all, reliability does not imply infallibility): whenever, in the past, Neda* relied on their say-so, she was not disappointed. By any account of testimonial justification in the literature, in this variation of the case Neda is justified to believe vaccines are not safe: according to anti-reductionism, this is because she has no defeaters to this testimony; according to reductionists, this is because she has inductive evidence of the reliability of these testifiers (Leonard 2023).

If Neda* is justified to believe vaccines are not safe, then she has a (in this case misleading) rebutting defeater for the scientific testimony that vaccines are safe. The defeater need not be a full defeater: laymen testimony might not be heavy enough – epistemically – to outweigh expert testimony. But Neda* will have reason to lower her confidence in the safety of vaccines: her (partial) rejection of scientific evidence is epistemically justified.

This is a case of misleading defeat. Of course, defeat to scientific testimony, generating epistemically permissible evidence rejection, can also be non-misleading: consider a case in which vaccinating toddlers is recommended by the experts to the sole benefit of the population at large (for generating herd immunity), since toddlers are not vulnerable to the virus

that the vaccine targets. At the same time, say that the vaccine is shown to have some side effects – albeit in very rare cases – the cause of which remains under-researched due to lack of funding: since these cases are rare, there is little incentive to invest in identifying the cause of the problem. Furthermore, say that Neda* is well aware of all of these facts, and thus she rejects scientific testimony that the vaccine is safe for her toddler and decides not to vaccinate him. This is a standard case of non-misleading rebutting defeat: Neda* is not only justified to reject the expert testimony that the vaccine is safe for her toddler; she is also, arguably, morally right to do so.

Justified evidence rejection need not only come through evidence against the proposition at stake (i.e. rebutting defeat). It can come about – and most often, I believe, it does come about – from undercutting defeat: reason to believe the expert source is not trustworthy. Consider again vaccine scepticism: sociological studies investigating vaccine hesitancy in Black and Caribbean communities in the UK, for instance, suggest that distrust in the safety of vaccines ultimately boils down to distrust of the National Health Service and medical science (Adekola et al. 2022). The thought is, in a nutshell, that a solid inductive basis suggests that the interests of these communities are not forefront concerns of these actors: historically, for instance, new medicines are not often tested on Black subjects before being commercialised. If so, this inductive evidence consti-tutes itself in undercutting defeat to the expert testimony in question. And, again, while undercutting defeat is often misleading when it comes to scientific expert testimony, it need not be such.

The above are ways in one can epistemically justifiably (partially or fully) reject evidence from highly reliable sources. These instances (i.e. instances of justified evidence rejection) will not make up the subject of this book. Likely, though, these will be the most ubiquitous instances on the ground: we are highly reliable cognitive machines. Bracketing very isolated cases of biased and heuristics-based cognition (which are often biological adaptations themselves), we are very good at responding to our epistemic environment: one can see this from the fantastic practical successes we enjoy as a species, which would not be possible without the associated epistemic high performance.

1.3.2   Evidence Resistance

Evidence resistance is an oddball in our species' cognitive life. As I will argue, it is an instance of epistemic malfunction of our cognitive system – similar to other input-level malfunctions occurring in other biological traits.

On a first approximation, evidence resistance can occur either in virtue of doxastic defeat or independently of it. Doxastic defeat (also sometimes referred to as psychological defeat in the literature) is defeat that lacks epistemic normative power but induces belief loss or downwards confidence adjustment nevertheless. The paradigmatic case of this has to do with proper updating on unjustified priors: I unjustifiably believe that all vaccines are unsafe and update accordingly to 'the COVID vaccine is not safe'. Some equate proper updating with rationality, in virtue of the epistemic value of coherence; most, however, shy away from offering such epistemic praise to cognisers who are fully coherent but completely disconnected from reality: take the perfectly coherent Nazi, for instance. Are we comfortable to call them perfectly rational? I would personally prefer to assign positive evaluative properties to a slightly incoherent version thereof – on both epistemic and moral grounds.

As the reader might have already guessed, my preference lies squarely with the second camp – the one that doesn't attribute much epistemic value to coherence alone and thus is sceptical about taking proper updating to be the mark of rationality. Not much will hinge on this for the rest of this book though. If impatient to read the relevant discussion, the reader can skip to Chapter 11.

Importantly, doxastic defeat need not occur via proper updating: improper updating is also an option (i.e. giving extant priors more evidential weight than they would deserve, even were they to be justified). Anchoring bias in all of its incarnations is a paradigmatic case.

Finally, evidence resistance need not be the result of updating at all – be it proper or improper. One such non-doxastically sourced, less common, and most simple variety can be an unexplained one-off instance of evidence resistance: maybe I'm looking straight at the table in front of me and, due to tiredness or lack of focus, I fail to notice the cup lying on it in plain view. Or say that I am very depressed and thus find it impossible to update on all of the evidence that my life is going really well.

Most commonly, though, non-doxastically sourced evidence resistance will be sourced in some variety of bias. Biases come in various shapes, and they can present as cognitive ('cold') biases (e.g. mental noise, heuristics) or motivational ('hot') biases (e.g. wishful thinking). To be clear, in many instances this variety of evidence resistance will be biologically beneficial, evolved in virtue of its biological benefits, and thus arguably practically rational. Compatibly, though, biased reasoning is epistemically deficient reasoning. Testimonial injustice is a paradigmatic case of evidence resistance due to bias: the hearer fails to give the testifier the level of credibility

that she deserves in virtue of a sexist bias that leads them to downgrade them as a testifier.

1.4   Conclusion

This chapter has done two main things: first, I looked at some of the recent literature on evidence resistance in social psychology, and I have argued that it misses important epistemological distinctions – such as, crucially, the distinction between epistemically justified evidence rejection and epistemically impermissible evidence resistance. I have then put forth a taxonomy of evidence rejection to help with isolating the problematic instances thereof, which will be my concern in the remainder of this book. In the following chapters, I will zoom in on evidence resistance and investigate the epistemological resources we need in order to explain its epistemic impermissibility.