

Concluding Remarks

The Way Forward in Policy and Practice^{*}

Disinformation is widespread and harmful, epistemically and practically. We are currently facing a global information crisis that the Secretary-General of the World Health Organization (WHO) has declared an ‘infodemic’.¹ Furthermore, crucially, there are two key facets to this crisis (i.e. two ways in which disinformation spreads societal ignorance): one concerns the widespread sharing of disinformation (e.g. fake cures, health superstitions, conspiracy theories, political propaganda, etc.), especially online and via social media, which contributes to dangerous and risky political and social behaviour. Separately, though at least as critical in the wider infodemic we face, is the prevalence of resistance to evidence: even when the relevant information available is reliably sourced and accurate, many information consumers fail to take it on board or otherwise resist or discredit it due to the rising lack of trust and scepticism generated by the polluted epistemic environment (i.e. by the ubiquity of disinformation). What we need, then, is an understanding of how to help build and sustain more resilient trust networks in the face of disinformation. To this effect, we need a better understanding of the nature and mechanisms of disinformation and of the triggers of evidence resistance.

Evidence Resistance

We have increasingly sophisticated ways of acquiring and communicating knowledge, but efforts to spread this knowledge often encounter resistance to evidence. Resistance to evidence consists in a disposition to reject evidence coming from highly reliable sources. This disposition deprives

^{*} From White Paper: ‘Trust, Disinformation, and Evidence Resistance’ (2022, with Chris Kelp). Submitted in response to the UK Parliament Call for Evidence: ‘Fighting Disinformation with Trustworthy Voices’. Published on the UK Parliament website here: <https://committees.parliament.uk/writtenevidence/111503/pdf/>.

¹ www.who.int/health-topics/infodemic#tab=tab_1.

us of knowledge and understanding and comes with dire practical consequences; recent high-stakes examples include climate change denial and vaccine scepticism.

Until very recently, the predominant hypothesis in social epistemology and social psychology principally explained evidence resistance with reference to politically motivated reasoning: on this view, 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. Typically, epistemologists who have explored the consequences of this empirical hypothesis take its merits at face value.

However, on closer and more 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 (often rationally justified) 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. Epistemologically, the worry is that the hypothesis is ineffective at making crucial distinctions among a number of phenomena, such as (1) concerning epistemic status: between irrational resistance to evidence and rationally justified evidence rejection; (2) concerning triggers: between instances of motivated reasoning on the one hand and epistemically deficient reasoning featuring cognitive biases and unjustified premise beliefs on the other; and (3) concerning strategies for addressing the phenomenon of evidence resistance: between targeting widely spread individual irrationality on the one hand and targeting an unhealthy epistemic environment on the other.

Furthermore, difficulties in answering the question as to what triggers resistance to evidence have a very significant negative impact on our prospects of identifying the best ways to address resistance to evidence. If resistance to evidence has one main source (e.g. 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 of social, emotive, and cognitive phenomena – we would have to develop much more complex interventions at both individual and societal levels.

My results suggest that the widespread irrationality hypothesis assumed by the politically motivated reasoning account of evidence resistance is

incorrect: humans are very reliable cognitive machines in spite of relatively isolated instances of biased cognitive processing or heuristics-based reasoning. Irrational resistance to evidence is rare and is an instance of input-level epistemic malfunctioning that is often encountered in biological traits the proper function of which is input-dependent: just like our respiratory systems are biologically malfunctioning when they fail to take up easily available oxygen from the environment, our cognitive systems are epistemically malfunctioning when they fail to take up easily available evidence from the environment.

What is often encountered in the population, however, is rationally justified evidence rejection due to overwhelming (misleading) evidence present in the (epistemically polluted) environment of the agent. When agents rationally reject reliable scientific testimony, they often do so in virtue of two types of epistemic phenomena: rebutting epistemic defeat and undercutting epistemic defeat. Rebutting epistemic defeat consists, often, in testimony from sources one is rational to trust that contradicts scientific testimony on the issue. These sources will be rationally trusted by the agent because of an excellent track record of testimony: they are overall reliable testifiers in the cognitive agent's community (which is why it is rational for the agent to trust them), but they are mistaken about the matter at hand. Reliability is not infallibility – it admits of failure.

One often-encountered trigger for rational evidence rejection is undercutting epistemic defeat. Undercutting epistemic defeat is evidence that suggests that a particular testimonial source is not trustworthy: relevant examples include misleading evidence against the reliability of a particular source of scientific testimony or a particular media outlet or against the trustworthiness of a particular public body. In vaccine-sceptic communities, for instance, we often encounter worries that the scientific community or the NHS does not have the relevant communities' interests in mind when they recommend vaccine uptake. These worries, in turn, are, once more, often rationally sourced in otherwise-reliable testimony (testifiers within the agent's community that the agent trusts due to their excellent track record but that are wrong on this particular occasion).²

These results, in turn, illuminate the best strategies to address the phenomenon of evidence resistance. Two major types of interventions are required:

² See, for example, *Scotland, African Voices: The Covid19 Vaccine Debate* (2022; documentary film, co-developed with Josephine Adekola, PI). Full documentary available here: <https://youtu.be/h1yNAZffpOg>.

- (1) For combatting rational evidence rejection: engineering enhanced social epistemic environments. This requires 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: (1.1) Quantitatively enhanced reliable evidence flow. This is a purely quantitative measure aimed at outweighing rebutting defeaters in the agent's environment. 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. (1.2) Qualitatively enhanced reliable evidence flow. This is a qualitative measure that aims to outweigh misleading evidence via evidence from sources that the agent trusts – that are trustworthy vis-à-vis the agent's environment (see below on context-variant trustworthiness). (1.3) Quantitatively and qualitatively enhanced evidence aimed at combatting undercutting defeat (misleading evidence against the trustworthiness of reliable sources). This involves flooding evidence-resistant communities with evidence from sources that they trust in favour of the reliability of sources they fail to trust due to misleading undercutting defeaters.
- (2) For combatting (relatively isolated) cases of irrational evidence resistance due uptake cognitive malfunction: increasing availability of cognitive flexibility training (e.g. in workplaces and schools, alongside anti-bias training; Chaby et al. 2019, Sassenberg et al. 2022). Cognitive flexibility training helps with enhancing open-mindedness to evidence that runs against one's held beliefs and with opening up alternative decision pathways.

Disinformation

My results show that disinformation need not come in the form of false content but rather consists of content with a disposition to generate ignorance under normal conditions in the context at stake. This predicts that disinformation is much more ubiquitous and harder to track than it is currently taken to be in policy and practice: mere fact-checkers just won't be able to adequately protect us against disinformation because disinforming does not require making false claims. Disinformation is ignorance-

generating content: content X is disinformation in a context C iff X is a content unit communicated at C that has a disposition to generate ignorance at C under normal conditions. The same communicated content will act differently depending on contextual factors such as the evidential backgrounds of the audience members, the shared presuppositions, extant social relations, and social norms. Generating ignorance can be done in a variety of ways – which means that disinformation will come in diverse incarnations, including false content, true content with false implicatures, false presuppositions, epistemic anxiety-inducing content, misleading evidence, and defeat.

What all of these ways of disinforming have in common is that they generate ignorance – either by generating false beliefs or by generating knowledge loss. Importantly, this capacity to generate ignorance will heavily depend on the audience's background evidence/knowledge. A signal r carries disinformation for an audience A wrt p iff A 's evidential probability that p conditional on r is less than A 's unconditional evidential probability that p and p is true.

Some of the best disinformation-detection tools at our disposal, targeting mainly false content, will fail to capture most types of disinformation. They are just the beginning of a much wider effort that is needed in order to capture disinformation in all of its facets rather than mere paradigmatic instances thereof, which involve false assertions. At a minimum, we need to build fact-checkers that track pragmatic deception mechanisms as well as evidential probability-lowering potentials against an assumed (common) evidential background of the audience.