**What is Knowledge Resistance?**

Kathrin Glüer & Åsa Wikforss

Forthcoming in *Knowledge Resistance in a High-Choice Information Environment*, Routledge, ed. J. Strömbäck.[1]



## 1. Introduction

Talk about knowledge (or fact) resistance is becoming increasingly common. A whole host of more or less related phenomena are labelled as such not only in research contexts, but also in public debate and social media discourse. The general impression seems to be of these "resistance phenomena" becoming more common and severe. To take just one example, there seem to be increasing numbers of people in the Western World who do not believe that vaccines are safe, despite strong and easily available evidence to the contrary. At the time of writing, the problematic nature of such resistance is becoming all too clear with the fourth wave of the Covid-19 pandemic sweeping around the world and bringing severe illness almost exclusively to the unvaccinated. While we have some indications of what kind of factors are driving vaccine skepticism, the general mechanisms of resisting knowledge aren't well understood yet. Progress on this front requires research of an unusually high degree of interdisciplinarity. To be successful, interdisciplinary research always requires conceptual calibration. In this case, the situation is aggravated both by unusually many disciplines having to talk to one another and the initial problem description.

[1] The paper is based on work carried out in close collaboration with Levi Spectre, within the cross-disciplinary research program *Knowledge Resistance: Causes, Consequences and Cures*, funded by Riksbankens Jubileumsfond.

As commonly used, talk of knowledge resistance tends to remain too loose and too metaphorical for gaining a precise and useful understanding of the core phenomena and the mechanisms producing them. It is thus essential to be precise on what is to be investigated. Given the nature of the initial problem description, this will be to some extent a matter of definition. In what follows, we shall offer a philosophical guide to the concepts we take to be most important for any empirical investigation of knowledge resistance and work towards a sufficiently precise characterization of it.

Knowledge resistance involves ignorance, but not all ignorance is the result of knowledge resistance. While ordinary ignorance can for instance result from mere lack of information, it is distinctive of ignorance resulting from knowledge resistance that information or evidence has been resisted. And to be resisted, that information or evidence has to be available to the resister. As a first approximation, then, we can say that *knowledge resistance involves resisting available evidence*.

In what follows, we shall unpack this characterization in more detail. There are two things, however, that we would like to be upfront about from the very start. One is that to get at the core of the phenomena of interest, the resistance involved in knowledge resistance should be construed as a form of *irrationality*. When it comes to evidence, what is resisted is the support it provides for a claim or conclusion. Resisting available evidence thus involves not drawing *proper* conclusions: conclusions supported by the evidence. But support comes in degrees. A conclusion might get *some* support from the evidence, but not enough to outrightly believe it. Moreover, different parts of the available evidence can point in different directions. In that case, the question is which conclusions can properly be drawn on the basis of *all* of it.

Here is a little toy story illustrating this. Imagine you are a detective investigating a "manor murder". Someone has been murdered in a remote English country manor. You have found bloody fingerprints in the butler's pantry. This is evidence pointing in the direction of his being the murderer. A little later, you also find the murder weapon, a sharp knife missing from the kitchen. You know that the butler and the cook are great friends, and that the cook has a watertight alibi while the butler does not, so this too points towards the butler. But then, you learn that the lord of the manor can expect large financial benefit from the death of the victim. Moreover, his alibi turns out to be fake. At this point in your investigation, you weigh all the evidence you have gathered and conclude that while both a suspects, you cannot yet draw a conclusion as to who the murderer is – the butler or the lord. Clearly, this is wise. And it should not be classified as knowledge resistance – while you do have evidence that supports either conclusion, the support is not sufficiently strong.[2]

This illustrates that not every conclusion that is to some extent supported by some available evidence is a *proper* conclusion to draw. The proper conclusion is determined by all the (relevant) available evidence *together*. And it can be of various forms; it can be, for instance, that something is rather likely, but not certain, and it can also be that we have no idea yet (as to who the murderer is, for instance).

The other thing that we would like to bring out from the very start concerns that which is resisted: evidence. Evidence can take many forms – we shall return to that, but what we take to be relevant to the core of the resistance phenomena is *empirical* evidence, evidence

[2] Nor would it have been knowledge resistance to refrain from drawing the conclusion that it was the butler before you found out about the lord's motive and false alibi – on the contrary, that again would have been the wise thing as the evidence available to you at that time was just not sufficiently strong yet.

concerning things we can investigate and come to know, at least in principle, by empirical means. Such means include scientific experiments, archeological digs, the study of historical documents, surveys and opinion polls, as well as many other things including the use of our eyes and ears and other senses.

The notion of knowledge resistance we suggest to work with, then, construes knowledge resistance as a form of irrational resistance to the available empirical evidence. This characterization needs to be further unpacked to be of use. Both with respect to the notion of resistance and to that of available evidence, we have only scraped the surface so far. Nothing has been said, for instance, about the psychological mechanisms involved. Another thing we need to say more about is what evidence is and what it means for evidence to be available. In what follows we shall map out the conceptual terrain in more detail and locate some of the relevant, central questions and results from the empirical literature on this map. Sections 2 and 3 will be about the notions of knowledge and evidence. In section 4, we explore two things in tandem: the notion of rationality and the kind of psychological mechanism that might explain knowledge resistant belief formation. Motivated reasoning will be our example here. The final section quickly probes the idea of distinguishing between two types of knowledge resistance, wide and narrow, and points to some complications relating to it.

**2. Knowledge**

Although the concept of knowledge is a non-technical concept, playing a central role in our everyday lives, a meaningful discussion of the nature of knowledge resistance requires explicating the concept. To begin, knowledge resistance concerns what philosophers call propositional knowledge, i.e. knowledge *that* something is the case, as opposed to knowledge *how*, skills and abilities. Propositional knowledge involves the subject holding a proposition *p*

true: it involves a *belief*. If I know that global warming is caused by carbon dioxide emissions, then I believe that global warming is caused by carbon dioxide emissions. To determine whether someone is knowledge resistant, therefore, one has to determine what they believe. This is not as straightforward as it may seem, and there are well known experimental challenges here. For instance, opinion polls are notoriously unreliable guides to belief, especially in the context of politics. One challenge is the phenomenon of expressive belief or "cheer leading", where someone expresses a view they do not genuinely hold in order to signal group membership and allegiance. When polls after the American presidential election 2020 show a large portion of Republican voters (around 50-70 percent) accepting the statement that the election was stolen, it therefore does not follow that they all actually believe that the election was stolen. They may just cheer on their preferred leader in his campaign to sow doubt on the election results.[3]

Knowledge requires belief, but belief is of course not sufficient for having knowledge – for one thing, what is believed also needs to *be true*. I cannot know that MMR vaccines cause autism, no matter how strongly I believe this to be the case, since it is false that MMR vaccines cause autism. This, in turn, implies that one can only know that which *can* be true or false. Put in philosophical terms, to be knowable, a content has to be *truth evaluable*. This too causes complications since it is philosophically controversial exactly which of our statements or mental states have such contents. This holds in particular for value statements such as "Lying is wrong", "Economic inequality should be counteracted", or "Bach was a great

[3] For more on this, see for instance Bullock and Lenz 2019, Badger 2020. Another phenomenon that might be relevant here is *choice blindness*. In choice blindness experiments, a surprisingly large number of participants seem to have little or no commitment to what they claim to believe in opinion polls. Using manipulated forms, researchers in such experiments manage to convince around 50 percent of interviewees of having expressed a view very different from what they actually stated it to be. Moreover, these individuals subsequently are willing to provide arguments for the new view, and the attitude change (if that's what it is) seems to last over time (Strandberg et. al. 2018).

composer". According to some philosophers, expressions of value are not factual statements, but more akin to things like screaming "Ouch!" – they express emotions but do not describe the world. On a psychological level this would mean that value expressions do not express beliefs and therefore are not even candidates for being the result of knowledge resistance. On a metaphysical level, such expressivism amounts to rejecting the idea that there are any value facts, such as moral or aesthetic facts.

But even if there are value facts, the relation between *normative* or value judgments and other (non-normative or *descriptive*) statements is complicated, to say the least. If someone doubts that Bach indeed was a great composer, it might just be possible to argue for that claim on the basis of *purely descriptive* characteristics of his music. But if someone holds that killing a person is always morally wrong it is at least much harder to see how any empirical data or descriptive facts would even be relevant. Such a statement might express a normative truth that holds irrespective of all the empirical or descriptive facts. Consequently, even if value judgements have truth evaluable contents and express beliefs, it would still not be clear whether they are candidates for knowledge resistance. Whether or not there are value facts, that is, value judgements just might not be such that empirical evidence is relevant to them.

This is important to remember when assessing the empirical literature on knowledge resistance and various types of emotion driven attitudes, since some of the experiments concern policy attitudes and these always involve values. The complication is that such experiments invariably provide subjects with (made up) empirical data. But if evidence is irrelevant to value judgments, then it is perfectly possible to accept *all* the relevant known empirical facts and yet reject a proposed policy without being irrational. Thus, one might accept that stricter gun control laws decrease deadly violence and yet reject such laws

because one values the right to freely carry guns more than a decrease in deadly violence. This would not mean that empirical knowledge is *irrelevant* to policy decisions; given that we have certain goals (such as reducing deadly violence) we will always need empirical knowledge about how best to reach this goal. But the question whether this is a goal worth having might not be one to which empirical evidence is relevant.

Unfortunately, these questions concerning the nature of value judgements can seriously complicate the interpretation of experimental results. A famous example of such an experiment concerns people's attitudes towards the death penalty (Lord, Ross and Lepper 1979). When people with partisan attitudes towards the death penalty were shown a scientific article concerning the link between capital punishment and crime, this did not affect their attitudes: those who were against the death penalty were not moved by (made up) evidence that it decreases crime, while those who were in favor were not moved by evidence that it does not decrease crime. This could be because people did not draw proper conclusions from the presented evidence, and thus a result of knowledge resistance, but it could also be because their attitudes rest on value judgments that are not influenced by empirical statements.[4] If you hold that it is always wrong to kill a person, then you will be against the death penalty no matter the empirical evidence.

Some experimental work on policy attitudes suggests that people do treat policy questions as amenable to factual or empirical arguments (whether or not this is philosophically the right position to take). Thus, in Taber & Lodge 2006, US college students were first asked for their

[4] There is a twist here, since the initial attitudes of the participants in the experiments actually were *strengthened* when they were presented with counter-attitudinal evidence. However, this is plausibly explained not as a (perverse) inference but as an emotional reaction to attempts to dislodge their deeply held attitude.

attitudes on affirmative action and gun control. Then, they studied or rated the strength of various arguments for and against these policies. Among other things, the study provides support for what Taber and Lodge call *disconfirmation bias*: people actively seek out attitudinally incongruent evidence in order to counterargue it. Moreover, this was more pronounced in politically knowledgeable subjects. It should be noted though, that the participants were told that they would have to present the debate in an objective manner to other students and they were explicitly asked to concentrate on what made the arguments weak or strong, and to leave their feelings aside, all of which would have primed them to think of the policy questions as empirical or factual questions. Indeed, as Taber and Lodge note, despite these instructions a good number of the participants made "simple content-free affective statements… to the effect 'I like (don't like) this argument or conclusion" (2006, p. 763).

Given these complications, experiments on knowledge resistance should be focused on judgments that are clearly factual, and whose truth can be investigated by scientific methods. Moreover, it is advisable to stay away from matters of fact where there is genuine uncertainty about the truth of the statement, and stick to cases where there is expert agreement. This, largely, is also how experimental research on knowledge resistance has been carried out. No doubt, certainty is hard to reach, and on occasion there has been scientific consensus on something that later turned out to be false, but these cases are the exception and the difficulties separating that which we have very strong evidence to believe is true from that which is not should not be exaggerated – indeed, exaggerating these difficulties is precisely one of the tools used by those who want to stop people from accepting inconvenient truths, for instance about smoking causing cancer or about climate change.[5] Nor should complex

[5] See Oreskes 2010.

questions be avoided; being complex might make a question initially harder to answer but has nothing to do with whether it is empirical or not. Moreover, many of the questions concerning which there appears to be resistance are of precisely this sort: they are complex, but entirely empirical. After listing some of these – does burning fossil fuels contribute to global warming? Does permitting citizens to carry concealed weapon increase homicide rates? Does vaccination against the human papilloma virus lead to more unprotected sex? – and stressing their empirical nature, Dan Kahan thus points out that "[i]ntense and often rancorous conflict on these issues persists despite the availability of compelling and widely accessible empirical evidence" (2017, 55; cf. also 2016, 1f).

Not everyone seems to agree. Mikael Klintman (2019) uses considerations of complexity to argue that we need to distinguish knowledge resistance from fact resistance. According to Klintman talking about fact resistance implies that we are concerned with a simple black-and-white phenomenon -- clear and indisputable fact -- whereas to speak in terms of knowledge resistance is to conceive of the phenomenon as multifaceted, more profound and less categorical (2019, pp. 17-20). As an example of fact resistance, he mentions resistance to unambiguous facts about the prevention of HIV, as an example of knowledge resistance beliefs about the risks of genetic engineering and about immigration and crime. But to support the distinction, he does not appeal to complexity, but rather to the idea that knowledge is more than just facts, that “knowledge comes in different shapes and colours: through systematic laboratory experiments and modelling, but also through daily experiences” (2019, p. 23). From a philosophical point of view, however, it is hard to see the point of this. No doubt, we need to distinguish facts, states of the world, from knowledge of these facts. Knowledge requires true beliefs, facts do not, and there are countless facts that we will never have knowledge of. But this has nothing to do with complexity, it holds for

simple questions (or facts) as well as for complex ones.[6] Similarly, evidence clearly can come from very different sources, but that by itself amounts to no more than a distinction between precisely that: sources of evidence. It provides no reason for thinking that we cannot usefully identify a general type of phenomenon here: knowledge resistance as precisely resistance to available empirical evidence – regardless of whether this evidence is supplied by laboratory experiments or some other method.

This takes us to a third necessary condition on knowledge, in addition to the two identified above: *justification*. As stressed by philosophers since Plato, we need to distinguish between a lucky guess and knowledge, and the distinction hinges on the idea that when someone knows that *p*, then the belief that *p* is not only true but also justified, based on evidence or good reasons.[7] The resistance part of knowledge resistance concerns this third component, more precisely, it concerns how people respond to evidence. Justification requires the drawing of proper conclusions from the available evidence. Resisting it thus prevents us from acquiring knowledge – belief formed in knowledge resistant ways is not justified and thus cannot be knowledge.

## 3. Evidence

[6] No doubt, some facts are fairly simple (facts about what there is in my fridge, for instance) whereas others are incredibly complex (such as facts about the causes of global warming). But there are no sharp distinctions to be drawn here, and the history of philosophy is littered with abandoned attempts at separating simple facts from complex ones.

[7] Since a famous paper by Edmund Gettier in 1963 it's been much debated whether these necessary conditions are also sufficient. Gettier showed that there may be situations in which one has a true, justified belief but where intuition tells us that the belief does not qualify as knowledge since the justification in question is only accidentally related to the truth of the belief (as when one happens to look at a clock at 11.35 that in fact has stopped at 11.35 but which one has every reason to think is working). However, discussing the nature of knowledge resistance does not require settling the issue of whether the three conditions are jointly sufficient, it is enough that they are necessary.

In ordinary parlance, ‘evidence’ is used in a variety of ways. For instance, we may describe objects and events as evidence (for instance, bloody fingerprints, murder weapons, smoking guns or explosions). However, in the context of knowledge resistance it is useful to construe evidence as *propositional*, not objectual. What we are interested in are relations between two propositions, for instance *that there are bloody fingerprints in the butler's pantry* and *that the butler is the murderer*. The first provides "support" for the second, support of a kind often characterized as *inferential support*: if you believe it you have (a certain degree of) support, or justification, for inferring the second proposition from the first. Since such an inference is not a deductive inference, and the truth of the first proposition does not guarantee that of the second, this notion of the support empirical evidence provides is quite naturally understood in terms of probabilification: if the evidence-proposition is true, then it is more likely that the proposition it is evidence for is true, too.[8]

In the case of human knowledge there are two fundamental forms of empirical evidence. First, there is experiential evidence, provided by our senses, such as visual, auditory or tactual evidence. How do I know that there is a tree in my garden? I see it. How do I know that a plane just flew by? I heard it. Experiential evidence plays a central role in how we navigate our daily lives, but what is distinctive of human knowledge is the extent to which we also rely on a different type of evidence, provided by the testimony of other people. How do I know what you did yesterday? You told me. How do I know that Biden is the current president of the US? I read it in the New York Times. How do I know that global warming is caused by carbon emissions? I listened to a climate researcher. When it comes to scientific knowledge,

[8] Thinking of evidential support in terms of probabilification is a natural and very prominent way of understanding the notion with respect to empirical evidence, and we shall adopt this understanding here. For a survey, see Kelly 2016. The basic idea is that one (empirical) proposition *p* provides evidential support for another empirical proposition *q* if the truth of *p* makes that of *q* more likely – that is, more likely than *q* would have been independently of *p*. There are intricate questions concerning problems such that the problem of "old evidence" in the vicinity, but we can abstract from those here.

for instance, we largely acquire this through testimony (even the scientists themselves do their work on the basis of a large body of knowledge acquired through testimony). The same holds for knowledge about current events and society, knowledge of the sort that is relevant to our tasks as voters in a democracy – it is largely testimonial, and it is to a large extent provided by media of various sorts.

An important consequence of this is that human knowledge relies on *trust*. If I do not trust what you say I won't believe it, and unless I believe it, I won't acquire the knowledge that you make available to me. The central role of trust in human knowledge makes for a certain type of vulnerability, relating to misplaced trust and distrust. Distrusting reliable sources means missing out on available knowledge. The type of ignorance that ensues is lack of belief about a certain matter. If, instead, the trust is misplaced and one trusts unreliable sources, the resulting ignorance involves having a false and/or unjustified belief about the relevant matter.[9] A central question, therefore, is how to determine the trustworthiness of a source. In a high-choice information environment, where there is great variation in reliability, the question becomes more acute and more challenging, demanding more of the individual than in an environment where the choices are fewer and the sources more reliable. A particular challenge concerns expert testimony, since it is in the nature of expert knowledge that non-experts often have difficulties making an independent evaluation of the plausibility of what the experts say. If you tell me that there is an elephant in the living room, I am able to determine that I probably should not trust your statement (not on this occasion at least), but when it comes to scientific statements my strategy will have to be more indirect, relying on

[9] Kuklinski et.al. (2000) distinguish between being uninformed (lacking the belief that p) and misinformed (wrongly believing p), but this distinction is not completely apt in this context. The person who, because of low trust, does not update her beliefs on evidence made available to her is not well described as being uninformed (after all, she has been informed by being provided with the evidence) even though the end result is the same type of ignorance – a lack of belief in a certain proposition.

institutional criteria of expertise (for instance, I could investigate the professional qualifications of the person). Also, in the case of disciplines where there is a direct connection with technology, these produce many statements that can be evaluated by laypersons (airplanes fly, cars drive, computers calculate, medicines heal), which provide a form of indirect test of the veracity of the science.[10] Nevertheless, trust in expert knowledge is particularly vulnerable since it can be undermined simply by sowing doubt about the messenger, without any independent means of checking the message. In the light of this, it is unsurprising that the reason most often cited in support of different types of science denial (be it about vaccines, the climate, or GMOs) is a certain type of conspiracy theory about the scientists (Lewandowsky et.al. 2013).

Knowledge resistance mostly appears to involve resistance to testimonial evidence, not experiential evidence. One reason for this is that it often seems to involve politically charged issues where relevant knowledge – such as scientific knowledge or knowledge about society – is typically acquired on the basis of testimony. Our focus will therefore be on testimonial evidence. [11]

As we said above, we take the notion of evidence relevant to the resistance phenomena to be propositional. This means that evidential relations hold among propositions. We also suggested to adopt the quite natural understanding of the evidential support provided by empirical evidence in terms of probabilification. For these reasons, it is important to say

[10] For a discussion see Baurman 2007, pp. 153-157. Philosophers distinguish between esoteric and exoteric statements, where the former belong to the sphere of expertise and are more difficult for non-experts to evaluate and the latter are more comprehensible also to lay persons (Goldman 2001). Notice that the pragmatic tests of scientific truth are not completely reliable since a machine may work even if the underlying scientific theory is partly incorrect (a point in case is Newtonian physics).

[11] Some experimental work suggests that there might be resistance even to sensory evidence; see e.g. Kahan et.al. 2012; Ripberger et.al. 2017.

something about the kind, or *form*, of proposition relevant here. Can we say anything general about what it is the various sources of empirical evidence "say"?

There is of course a huge variety of sources providing testimonial evidence. The kinds of things they can "say" (be it by means of speech, text, or image) and thereby provide evidence for are, if anything, even more varied. Nevertheless, we can think about the form testimonial evidence takes via the structure of the justification it provides.[12] If the source is knowledgeable and trustworthy a recipient can acquire knowledge from the source. This is because that they are saying that *p* is a good indicator of *p*'s truth. But if the source is unreliable, this is not the case. The evidence therefore needs to be of a form that makes its evidential power dependent on the source's reliability: if it is true and the source is reliable, the truth of what the source says is more likely, but if the source is unreliable, it is not (or at least less so). To capture this structure, we suggest construing testimonial evidence as being of the form *S says that p*, where *S* can be any source that provides information – be it via speech, text, or images. On this construal, *what a source says* is not to be identified with the evidence – rather, what they say is what they provide evidence *for* by saying it (if sufficiently reliable). Thus, a reliable scientist NN's assertion that global warming is caused by burning fossil fuels provides testimonial evidence with the content *NN says that global warming is caused by burning fossil fuels*. What is it is evidence for is *that global warming is caused by burning fossil fuels*.

When it comes to epistemic justification, however, we are not simply interested in relations between propositions, but also in people's *relation* to the evidence, which is to say that we

[12] Our suggestion here, and the motivation for it, are modelled on Glüer's account of the content and justificatory role of perceptual experiences (cf. e.g. Glüer 2009, 2016). Resistance to sensory evidence, if any, could usefully be modeled on this account, too.

are interested in their psychological states. To go back to our murder mystery: Imagine that you in the meantime have learned that the lord faked his alibi because he was with his lover and that DNA traces from the butler have been found under the victim's fingernails. Now you have much stronger evidence for the belief that the butler did it. At this point, your long-time partner falls ill and you get new assistant, DC Sally Sleuth. Sleuth hasn't seen, or heard of, any of the evidence you have uncovered, but she believes that the butler did it simply because she dislikes him. *Her* belief is just as true as yours, but if Sleuth indeed formed it without *having* any evidence, then her belief is not justified.[13]

How does this play out in the context of knowledge resistance? Does knowledge resistance only concern resistance to evidence that the subject *has*? Clearly, it cannot be that we are knowledge resistant simply because we are unaware of evidence. In that case every one of us

[13] Of course, you then tell Sleuth what you have found. And naturally enough, she still believes that the butler did it. She might even say: "I knew it!" Of course, *that* would not be true if she initially formed it without having any evidence. Assume that she initially was not justified in forming the belief. A more tricky question then is whether she *now* is justified. Philosophers distinguish between what you have justification for believing (whether you actually believe it or not – this is often, somewhat misleadingly, called "propositional justification") and being justified in what you actually believe ("doxastic justification"). Whether you have doxastic justification for a belief is often taken to depend on whether you hold it *based* on good reasons or evidence. In Sleuth's case, this is not clear from what we have said about her. We assume that she initially formed the belief that the butler did it without having any evidence. But whether she *now* is basing her belief on the evidence you have given her would seem to depend on what would happen if it turned out that the evidence isn't as strong as it seems now (you might for instance learn that, shortly before the murder, the victim dragged the half-drowned butler from the pond with his bare hands). Arguably, Sleuth's present belief is justified only if she would give it up (or at least lower her confidence in it) in such a situation.

would be radically knowledge resistant all the time. At the same time, requiring possession of the evidence in the sense of belief might be too strong: Some of the pertinent cases seem to involve *avoidance* of evidence. If I know that I can find out whether the butler did it by looking in his pantry, but I decide not to look because I do not want to believe he did it, then I'm plausibly resisting knowledge, even if my belief formation as such is not irrational. Similarly, if I suspect that Trump has done some not so great things, and I avoid reading the New York Times since I think they write about this, then I'm plausibly resisting knowledge. In order to characterize these cases as a form of knowledge resistance, we will need to understand the notion of available evidence to not only include all the evidence a subject has, but to be wider than that. In particular, we will need to say more about what it means to avoid available evidence. We shall return to this, rather difficult, question below. First, however, we shall focus on cases where the subject has the evidence.

**4. Rationality, irrationality, and motivated reasoning**

Knowledge resistance involves an irrational response to evidence. But we are not interested in just any instance or pattern of irrational belief formation. We are not interested in occasional miscalculations, for instance. Nor in systematic ones due to nothing but cognitive limitation. What we are interested in are patterns of irrationality that are the result of certain kinds of psychological mechanisms, mechanisms different from rational information processing in ways that invite describing these mechanisms in terms of resistance. A prime candidate for such a mechanism is what psychologists call *motivated reasoning*.

Motivated reasoning is belief formation improperly driven by what philosophers call desire: any kind of wish, want, urge, hope or other "pro-attitude" regarding a proposition's truth. That is, it is belief formation driven by desire *rather than* by epistemic reasons. Motivated

reasoning has been investigated by social psychologists since the 1950's (for an overview, see Kunda 1990). A common description of the phenomenon is that it involves believing what one *wants* to believe, rather than what one has good reasons to believe. However, this description is not unproblematic since it construes belief as a kind of action, one that is under the control of the will, a position philosophers call *doxastic voluntarism*. There are strong empirical reasons to believe that doxastic voluntarism is false and that the kind of control we have over our beliefs is, at most, indirect (see Alston REF?). I can no more get myself to believe that global warming is a hoax than I can get myself to believe that 2 plus 2 equals 5 – no matter how much money you offer to pay me for forming either of those beliefs. (I could, however, try to totally immerse myself in the world of climate deniers and perhaps that would eventually lead me to become a skeptic). Indeed, as has been noted in the literature on motivated reasoning, people do employ various indirect strategies to protect desired beliefs, strategies utilizing the fact that the best way to influence beliefs is through arguments and evidence. Ziva Kunda writes: "I propose that people motivated to arrive at a particular conclusion attempt to be rational and to construct a justification of their desired conclusion that would persuade a dispassionate observer" (1990, p. 84). A useful, less voluntaristic characterization of motivated reasoning is provided by Kahan, who defines it as "the tendency of individuals to unconsciously conform their assessment of information to some goal collateral to assessing its truth" (2016, 2). These collateral goals are often called "directional goals". And it is important here that the tendency to conform one's judgement to a directional goal is supposed to be *caused and explained* by having the directional goal.

Understanding the irrationality of motivated reasoning relevant to knowledge resistance requires reflecting on how it relates to the rationality of action. On the standard construal, the rationality of action depends on two components: belief and desire (or pro-attitude). If I want another beer, and I believe there is beer in the fridge, then I have a reason to go to the fridge.

(Of course, such a reason might be overridden by others – I might for instance plan to drive later.) This type of rationality is sometimes called *practical* rationality, since it leads to action (or at least to intentions to act), and reasons for action are called practical reasons. It's been much discussed in philosophy whether there can be practical reasons for belief, whether it can ever be rational to form a belief because one desires to do so. According to some philosophers, beliefs that are formed on the basis of practical reasons are always irrational (Smith 2004), according to others it can be perfectly rational to form a belief if, for instance, an evil demon threatens one's life unless one forms the belief. Even if one accepts that belief can be practically rational, however, it is important to distinguish practical rationality from *theoretical or epistemic* rationality. Theoretical rationality concerns the epistemic basis of belief, the extent to which it is based on evidence, and wishful thinking will always be epistemically irrational – regardless of whether it could be described as practically rational. The distinction is of some importance to the discussion of motivated reasoning. Thus, Kahan has suggested that it can be rational for an individual to hold on to beliefs that have become symbols of membership in what he calls "identity-defining affinity groups", groups central to an individual's social identity. According to Kahan, this can be rational even in the face of strong counterevidence, because for someone getting "the 'wrong answer' in relation to the one that is expected of members of his or her affinity group, the impact could be devastating: loss of trust among peers, stigmatization within his community, and even loss of economic opportunities" (Kahan et. al. 2017, 57). Thus, for conservatives in certain parts of the US, it may be important to endorse climate skepticism. But even if one accepts the idea that motivated belief can be practically rational, it is important to stress that in the epistemic

sense, it would nevertheless be irrational. And that is what matters from the point of view of knowledge resistance.[14]

The relevant notion of rationality, therefore, is epistemic rationality. This, in turn, can be understood according to different models. The most commonly used is Bayesianism. Bayesianism is a simple and natural way of using the idea that empirical evidence can be understood in terms of probabilification to model the epistemic rationality of updating a belief system on the basis of new information.[15] Bayesianism models a belief system as a space of propositions over which a probability function distributes its values. These values, conventionally set to range between 0 and 1, model a subject's credences or "subjective probabilities", i.e. the degrees to which they hold the propositions true. These degrees can be very high (1 - outright belief) or very low (0 - outright disbelief), but the subject might also find some things more or less likely, without being certain. Using a probability function to model credences places rationality or coherence constraints on them – such functions for instance require that the credences for p and not-p add up to 1. The function also assigns conditional probabilities: probabilities of one thing being true given the truth of some (other) thing. These answer questions like: How likely is it that there is fire if there is smoke? How

[14] It is sometimes suggested that all reasoning is motivated, and that the difference between the good and the bad case is that in the good case the motive is accuracy, whereas in the bad case the motive is a different one, such as the desire to fit in with the group (see for instance Kunda 1990). Here, we will set this idea aside, but it is worth pointing out that while empirical research does show that accuracy improves when we are motivated to pay attention it does not follow that the reasoning itself is motivated – all that follows is that we perform better when we have a motivation to do so.

[15] Bayesianism is often criticized as unrealistic and overly idealized. Here, we need to distinguish between using Bayesianism as a model of actual information processing in the human brain or as a model of what rational processing would look like. But even if we are concerned with modeling actual brain processes, it is important that to be a useful model here it is not required that human subjects consciously run through the inferences required by Bayesian updating or even know anything about the probability calculus underwriting it. All that is required is that actual processing is in sufficient conformity to the patterns of processing predicted by Bayesianism. As far as we can tell, the jury is still out on whether it does. It is noteworthy, however, that many purported counterexamples have been successfully shown to be amenable to Bayesian treatments.

likely is it that the butler is the murderer if there are bloody fingerprints in his pantry? Or: how likely is it that consuming GMOs is detrimental to your health if a certain new age guru says that it is? Now, assuming that we have our probability function $P_o$ in place for a given belief system, we can use it to model how to rationally update that system in the light of new information. Updating will result in a new probability function $P_n$ that models the updated system. For getting from $P_o$ to $P_n$, Bayesianism uses a rule called *Bayesian conditionalization*. It tells you that when updating on new information *p* the updated credence or probability assigned to a proposition *q* should be identical to the old conditional probability of *q* given *p*:[16]

(BC) $P_n(q) = P_o(q|p)$

To illustrate: At the very beginning of case of the manor murder, you may think that it is as likely as not that it was the butler, but would go up to, say, 70% should you find bloody fingerprints in his pantry. Then, you do find bloody fingerprints in the pantry. Bayesian conditionalization tells you to now update your credence in the butler's being the murderer from .5 to .7.[17]

As you can see, a lot here hinges on the credences you start your investigation with. These are called your *priors*. Given that you are an experienced detective and already have a long and relevant information history, you might come to the case thinking that in cases of this sort it more often than not is the butler. Your prior would then be larger than .5. And this might even be reasonable. But now think of Sally Sleuth again. She formed her belief that the

[16] The conditional probability pf p on q in turn is determined by Bayes' theorem:
(BT) $P(q|p) = P(q) \times P(p|q)/P(p)$
When updating your belief in q in the light of new evidence p, this tells you to combine (by multiplication) your prior belief in q with the likelihood of getting evidence p in case q is true, and then to normalize by dividing by the likelihood of getting that evidence regardless of whether q is true or not.

[17] For illustration's sake, we here abstract from all the other evidence relevant to the question of who the murderer is, evidence that might make it rational to place an even higher credence on the butler, but also a lower one. Something about such precise credences being unrealistic?

butler did it on the basis of her dislike for him. Which on the face of it sounds like a clear case of motivated reasoning. But there is a wrinkle. Imagine that Sleuth comes to the case with a high prior on its being the butler if she dislikes him. She might, for instance, think of her butler-dislikes as some sort of indicator of their being murderers. She might even have already investigated quite a number of manor murder cases where butlers were the culprits precisely when she found them unsympathetic. Upon meeting the present butler, she finds that she does dislike him and updates accordingly. In this case, Sleuth's belief formation, even though there might well be something improper about it, is *not* a case of motivated reasoning. Rather, she updates her beliefs in a way that not only is in accordance with Bayesian updating, but indeed exhibits a certain internal rationality. Given that she thinks that her butler dislikes in fact indicate whether they are murderers, it makes a certain sort of subjective sense that she comes to believe the present butler to be one. What might be irrational here, if anything, it is Sleuth's prior belief about her dislikes, not the way she updates.

This illustrates that using Bayesianism to model *subjective* probabilities, or credences, and how to update them in the light of new information provides us with a notion of rationality that, in a certain sense, is subjective, too: It first and foremost tells us how to rationally update a *given* belief system, i.e. a subject's *actual take* on the world – including on what in fact makes what more likely. And a given belief system might be more or less in line with how the world actually is. Thus, even though an update might initially strike us as simply unreasonable, it might not be unreasonable through and through given a subject's prior beliefs. And while those priors themselves might have been arrived at in improper ways, locating the irrationality, if any, would in many cases require more and deeper knowledge; in particular, we would need to know more about the subject's *information history*. In Sleuth's

case, it is unlikely that her belief about her dislikes has been formed in a proper way, but in many other, more realistic cases, it is far less easy to tell to what degree a subject's prior beliefs are justified by their information histories and, if not, where exactly the irrationality is located.

When it comes to motivated reasoning, what is important is that here, directional goals play an improper role in the updating process itself. Psychological research on the nature of motivated reasoning suggests that a variety of mechanisms are involved (see Kunda 1990, 84-102). It can be a matter of biased memory search, where people search for beliefs that support the desired conclusion, accessing only part of their relevant beliefs. Directional goals can also influence the speed of reasoning, lead to skewed weighing of evidence, biased selection of inference rules and statistical heuristics, or biased evaluation of the methodological aspects of scientific studies or of the quality of expertise. There are however well known experimental challenges when it comes to studying these processes. One central challenge is to *determine* whether the reasoning is indeed driven by directional goals or whether it can be explained in non-motivational terms, by appealing to prior beliefs and expectancies.[18] When it comes to resisting testimonial evidence a subject has, prior beliefs about, or other considerations relevant to, the *reliability of the source* of the evidence become particularly interesting. If I am not sensitive to source *S*'s telling me that *p*, this may indeed be because I'm engaged in motivated reasoning but it could also be because I believe *S* to be unreliable or because my prior beliefs give me reason to think that *p* must be false (which in turn will influence my belief in *S*'s reliability). Both of these beliefs might themselves have been formed in

[18] Kunda (1990) discusses this challenge and argues that several studies fail to provide unambiguous support for the role of motivation in producing biased reasoning since the biases could also have been due to the role of prior beliefs (p. 92, 97). For an excellent philosophical survey of the relevant issues and challenges, see Hahn & Harris 2014.

improper ways, of course, but again, the impropriety would in that case lie somewhere in my information history, not in the particular updating process under investigation.

When it comes to the psychological sources of motivated reasoning in general, one can desire to protect, or resist, a belief as a result of any number of psychological factors – be it emotions such as fear, character traits such as vanity, or a wish to fit in with the group. For core cases of knowledge resistance, however, certain types of sources appear to be typical. In a discussion of science denial, Hornsey and Fielding (2017) suggest there are six different types of sources, what they call "attitude roots", underlying motivated reasoning. Among these are vested interests, fears and phobias, personal identity expression and social identity needs.

Especially social identity needs have come to play a prominent role in research on a form of motivated reasoning of particular interest: *politically* motivated reasoning, often considered to be the main explanation for the phenomenon of *fact polarization* (cf. esp. Kahan 2016a, b). In Kahan's words, fact polarization is "intense, persistent partisan contestation over facts that admit of scientific evidence" (Kahan 2016b, 1). In cases where the relevant scientific evidence is clear, univocal, and readily available, fact polarization appears to be a prime location at which to find knowledge resistance. We have already met Kahan's influential idea that subjects' identity protection needs motivate them to resist evidence against beliefs that have become symbols of membership in identity-defining social groups. For partisans, such groups are political, but the relevant beliefs are factual; they are beliefs admitting of scientific evidence, beliefs about the causes of global warming, vaccine safety, or the effects of GMOs on human health, for instance. Kahan and his group not only have carried out a number of important experimental studies investigating people's reactions to evidence regarding various such issues, they have also carefully calibrated their experimental designs to the detection of

political motivation, in particular as distinguished from confirmation bias (cf. Kahan 2016). Nevertheless, not even Kahan's designs would seem to escape an important confound.

Tappin et. al (2020a, b) point out that experiments for the detection of politically motivated reasoning typically are of one of two kinds. What is relevant here are those they call "outcome switching" designs. Such experiments use subjects' evaluation or endorsement of new information as their outcome variable. The subjects are randomly assigned one of two pieces of information, the substantive detail of which is held constant across conditions. But the implications this information has for subjects' political identities varies between conditions: In one condition, the information is identity-incongruent, in the other it is congruent. "The key result is that subjects' evaluation of the information differs by condition, and, in particular, that this difference is correlated with their political identities or preferences" (Tappin et.al. 2020a, 82). But people's political identities are often also correlated with their *prior beliefs* about the relevant issue. Switching the treatment information to render it incongruent with political identity and motivation *also* renders it incongruent with prior belief. And, as we illustrated above, prior belief can influence people's reasoning in the absence of political motivation. This does *not* mean that what is observed in outcome switching experiments in fact is not politically motivated reasoning. But it does mean that experiments of this kind do not allow us to tell it apart from reasoning explained by prior belief: "[T]he effect of prior belief (…) confounds inferences of political motivation in this design" (Tappin et.al. 2020a, 83). Ditto et.al. call this "an empirical catch-22 at the heart of all research on motivated reasoning" (2018, 285).

Even experiments in which subjects appear to adjust their estimate of the reliability of a source of evidence in response to whether what the source says is in line with their prior beliefs can be confounded in this way. As we said above, what someone says can be an

indicator of their reliability. If you, for instance, are presented with a source described as having all the credentials of an expert climate scientist, but also as saying that global warming is a hoax, you have very good reason to doubt their reliability. Doing so both appears reasonable and is fully in line with Baysianism (cf. Hahn & Harris 2014, Tappin et. al. 2020b). All in all, it seems fair to say that the detection of politically motivated reasoning faces important challenges, challenges that have not entirely been resolved to date.

**5. A wider notion of knowledge resistance?**

At the same time, the difficulties of disentangling the effects of motivation from those of prior belief raise the question of how wide the notion of knowledge resistance should be construed. There are at least two issues worth considering here, both of which have to do with a subject's information history. The first concerns updating on evidence the subject has but that is in fact explained by prior belief – where that prior belief itself results from irrational formation, in particular from misplaced trust. Such trust might be ungrounded, for instance because it is politically motivated.[19] We tend to think that cases of this kind are usefully included under the notion of knowledge resistance. However, we also tend to think that it might be useful distinguish between a narrow and a wider sense of knowledge resistance in order not to lose track of the difference.

The second issue concerns cases where the subject does *not* have the evidence because of certain features of their information environment.[20] As we already said above, we think that

[19] Baron & Jost (2019) stress the differences in media trust between Republicans and Democrats, and note that "four of the U.S. news/opinion sources that contained the highest proportion of false statements (the Rush Limbaugh Show, Glenn Beck Program, Fox News, and the Sean Hannity Show) were highly trusted by a strong majority (51%-88%) of 'consistent conservatives'" (2019, pp. 293-294). See also Cook & Lewandowsky 2016, p. 172, for a discussion of ungrounded trust.

[20] When it comes to political partisans, an important question concerns whether their political motivations might not only tend to correlate with prior belief, but also increasingly with more or less exclusive consumption of partisan media.

some cases of this kind are usefully counted as knowledge resistance. Care needs to be taken, however, to avoid including cases where people just have had the bad luck of due to no fault of their own ending up in highly unreliable information environments. Take for instance a person growing up in a fundamentalist sect. Such a person might develop a system of fundamentalist belief that is both subjectively rational and results from rational responses to all the evidence available to them (Baurman 2007 provides useful discussion and suggests how to model such a development). We might not want to count such a subject as knowledge resistant.

Fundamentalist conditions are rare, however. More often than not, subjects have quite a bit of control over which sources they get information from. Thus, ignorance might be the result of what media researchers call "selective exposure": active selection, in particular active avoidance of certain sources of information. We might therefore want to allow for a second kind of knowledge resistance in the wider sense: this kind involves resisting evidence a subject easily could have had but chose to actively avoid.

Exploring the wider notion of knowledge resistance further would take us too far afield, however. We would like to round of our considerations by just mentioning two of the issues in need of exploration here, both concerning news avoidance. One concerns irrationality: A subject's habits of news avoidance could for instance result in ignorance of the fact that CISA said that the 2020 US election was fair. This could in turn be the result of an ungrounded, epistemically irrational belief as to what sources can be trusted. But does the irrationality have to be epistemic? If my news avoidance results from nothing but my desire for entertainment, should the resulting ignorance nevertheless count as knowledge resistance?

We tend to think not, if only to keep the notion usefully sharp, but the question requires more discussion then we can provide here.

The other issue again concerns detection again. The challenge here is to separate failures to take in available evidence that are usefully described as resistance from those that derive from limitations such as limitations of time and attention. In a high-choice, high-traffic information environment there will inevitably be limits to what we can attend to. Ignorance resulting from these ought not to be counted as knowledge resistance, at least not automatically.[21]

[21] Indeed, a useful strategy to prevent people from taking up the available evidence, employed by authoritarian states, is precisely to flood the information channels with irrelevant information (cf. Wu 2018. See also chapter XX for discussion).

**References**


Alston, W. P. (1988). The Deontological Conception of Epistemic Justification. *Philosophical Perspectives*, *2*, 257–299. https://doi.org/10.2307/2214077

Badger, E. (2020, December 1). Most Republicans Say They Doubt the Election. How Many Really Mean It? *The New York Times*, 18.

Baurman, M. 2007. Rational Fundamentalism? An Explanatory Model of Fundamentalist Beliefs. *Episteme*, 4, 150-165.

Baron, J., & Jost, J. T. (2019). False Equivalence: Are Liberals and Conservatives in the United States Equally Biased? *Perspectives on Psychological Science*, *14*(2), 292–303. https://doi.org/10.1177/1745691618788876

Cook, J., & Lewandowsky, S. (2016). Rational Irrationality: Modeling Climate Change Belief Polarization Using Bayesian Networks. *Topics in Cognitive Science*, *8*(1), 160–179. https://doi.org/10.1111/tops.12186

Gettier, E. L. (1963). Is Justified True Belief Knowledge? *Analysis*, *23*(6), 121–123. https://doi.org/10.1093/analys/23.6.121

Glüer, K. (2009). In Defence of a Doxastic Account of Experience. *Mind & Language*, *24*(3), 297–327. https://doi.org/10.1111/j.1468-0017.2009.01364.x

Glüer, K. (2016). Intentionalism, Defeasibility, and Justification, Philosophical Studies, 173: 1007-1030.

Goldman, A. I. (2001). Experts: Which Ones Should You Trust? *Philosophy and Phenomenological Research*, *63*(1), 85–110. https://doi.org/10.2307/3071090

Hahn, U., & Harris, A. J. L. (2014). Chapter Two - What Does It Mean to be Biased: Motivated Reasoning and Rationality. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 61, pp. 41–102). Academic Press. https://doi.org/10.1016/B978-0-12-800283-4.00002-2

Kahan, D. M. (2016a). The Politically Motivated Reasoning Paradigm, Part 1: What Politically Motivated Reasoning Is and How to Measure It. In *Emerging Trends in the Social and Behavioral Sciences* (eds. R.A. Scott and S.M. Koslyn), (pp. 1-16) https://doi.org/10.1002/9781118900772.etrds0417

Kahan, D. M. (2016b). The Politically Motivated Reasoning Paradigm, Part 2: Unanswered questions. In *Emerging Trends in the Social and Behavioral Sciences* (eds. R.A. Scott and S.M. Koslyn) https://doi.org/10.1002/9781118900772.etrds0418

Kahan, D. M., Hoffman, D. A., Braman, D., & Evans, D. (2012). They Saw a Protest: Cognitive Illiberalism and the Speech-Conduct Distinction. *Stanford Law Review*, *64*(4), 851–906.

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, *1*(1), 54–86. https://doi.org/10.1017/bpp.2016.2

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, *2*(10), 732–735. https://doi.org/10.1038/nclimate1547

Kelly, T. (2016). Evidence. *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.) URL = <https://plato.stanford.edu/archives/win2016/entries/evidence/>.

Klintman, M. (2019). *Knowledge Resistance. How We Avoid Insights from Others*. Manchester University Press. https://www.manchesterhive.com/view/9781526158703/9781526158703.00005.xml

Kuklinski, J. H., Quirk, P. J., Jerit, J., Schwieder, D., & Rich, R. F. (2000). Misinformation and the Currency of Democratic Citizenship. *The Journal of Politics*, *62*(3), 790–816. https://doi.org/10.1111/0022-3816.00033

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Lewandowsky, S., Gignac, G. E., Oberauer, K. (2013). The Role of Conspiracist Ideation and Worldviews in Predicting Rejection of Science, PLoS ONE, 8 (10).

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*(11), 2098–2109. https://doi.org/10.1037/0022-3514.37.11.2098

Oreskes, N., Conway, E. M. (2010). *Merchants of Doubt. How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Climate Change*. Bloomsbury Press.

Ripberger, J. T., Jenkins-Smith, H. C., Silva, C. L., Carlson, D. E., Gupta, K., Carlson, N., & Dunlap, R. E. (2017). Bayesian versus politically motivated reasoning in human perception of climate anomalies. *Environmental Research Letters*, *12*(11), 114004. https://doi.org/10.1088/1748-9326/aa8cfc

Scott, R. A., & Kosslyn, S. M. (Eds.). (2015). *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource* (1st ed.). Wiley. https://doi.org/10.1002/9781118900772

Smith, M. (2003). Humeanism, Psychologism, and the Normative Story. *Philosophy and Phenomenological Research*, *67*(2), 460–467.

Strandberg, T., Sivén, D., Hall, L., Johansson, P., & Pärnamets, P. (2018). False beliefs and confabulation can lead to lasting changes in political attitudes. *Journal of Experimental Psychology: General*, *147*(9), 1382–1399. https://doi.org/10.1037/xge0000489

Taber, C. S., & Lodge, M. (2006). Motivated Skepticism in the Evaluation of Political Beliefs. *American Journal of Political Science*, *50*(3), 755–769. https://doi.org/10.1111/j.1540-5907.2006.00214.x

Tappin, B. M., Pennycook, G., & Rand, D. G. (2020a). Thinking clearly about causal inferences of politically motivated reasoning: Why paradigmatic study designs often undermine causal inference. *Current Opinion in Behavioral Sciences*, *34*, 81–87. https://doi.org/10.1016/j.cobeha.2020.01.003

Tappin, B. M., Pennycook, G., & Rand, D. G. (2020b). Rethinking the Link between Cognitive Sophistication and Politically Motivated Reasoning. *Journal of Experimental Psychology: General*. Advance online publication. http://dx.doi.org/10.1037/xge0000974

Wu, T. (2018). Is the First Amendment Obsolete? *Michigan Law Review*, *117*(3), 547–581. https://doi.org/10.36644/mlr.117.3.first