Review of Philosophy and Psychology
https://doi.org/10.1007/s13164-024-00737-w

Where is the Motivation in Motivated Numeracy?

Kathrin Glüer-Pagin[1] · **Levi Spectre**[2]

Accepted: 8 May 2024

Abstract
In a series of very influential papers, Dan Kahan argues for "the identity protective cognition thesis": the claim that politically motivated reasoning is a major factor explaining current levels of polarization over matters of fact, especially in the US. An important part of his case consists of experimental data supporting the claim that ideological polarization is more extreme amongst more numerate individuals. In this paper, we take a close look at how precisely this "numeracy effect" is supposed to come about. Working with Kahan's own notion of motivated reasoning, we reconstruct the mechanism that according to him produces the effect. Surprisingly, it turns out to involve plenty of motivation to reason, but no motivated reasoning. This undermines the support he takes the numeracy effect to provide for the identity protective cognition hypothesis.

Keywords Motivated reasoning · Fact polarization · Identity protection · Motivated numeracy · Knowledge resistance · Dan Kahan

That the more numerate or cognitively sophisticated also are "better" at motivated reasoning has lately become something of a commonplace. Science journalists, editorial writers, and the interested public, maybe unsurprisingly, revel in the seeming

✉ Kathrin Glüer-Pagin
kathrin.gluer@philosophy.su.se

✉ Levi Spectre
levisp@openu.ac.il

[1] Department of Philosophy, Stockholm University, Stockholm, Sweden

[2] Department of History, Philosophy and Judaic Studies, Department of Philosophy, The Open University of Israel, Israel, Ra'anana

Published online: 30 May 2024

irony of such "numeracy effects".[1] Describing Dan Kahan's influential findings, *New York Times* podcaster Ezra Klein for instance tells us: "The smarter the person is, the dumber politics can make them" (Klein 2020; p. 92). Numeracy effects certainly are interesting in their own right, but become even more so when claimed to be a key element in the case for the hypothesis that motivated reasoning is the source of the fact polarization that seems to be "a signature feature of contemporary political life" (Kahan 2016a, 1).

Fact polarization is "intense, persistent partisan contestation over facts that admit of scientific evidence" (Kahan 2016b; p. 1). And it indeed seems to be the case – in the US and maybe the Western World in general – that growing numbers of people either resist or embrace compelling and widely accessible scientific evidence along partisan lines. Such polarization matters when it concerns policy-relevant facts such as vaccine safety, the relation between gun control and violent crime, or that between human activity and climate change. It is therefore paramount to understand how it comes about.

In a series of highly influential papers, Kahan has argued that fact polarization results from politically motivated reasoning. More precisely, he has argued for "the identity-protective cognition thesis" (ICT), the hypothesis that politically motivated reasoning is identity-protective reasoning, and that this is a major driver of fact polarization (cf. e.g. Kahan et al. 2012; Kahan 2013; 2015; 2016a, b; Kahan et al. 2017). Kahan's case for ICT crucially depends on a numeracy effect – the effect is supposed to provide an essential piece of evidence supporting ICT over rival explanations.

Its being subject to cognitive sophistication effects should be intriguing to anyone interested in motivated reasoning, not least the philosopher. Understanding exactly how such effects are supposed to work is not easy, however, and there are quite different suggestions around. In this paper, we will try to make some progress here by offering a close reading of Kahan et al.'s work on "motivated numeracy". In Section 1, we'll spell out Kahan's notion of politically motivated reasoning, his explanation of fact polarization in terms of it, and the role of the numeracy effect in the case for ICT. In Section 2, we are going to reconstruct Kahan et al.'s understanding of the mechanism generating the numeracy effect. Surprisingly, this mechanism – we'll call it "Kahan's mechanism" – turns out to involve motivation to reason, but no motivated reasoning. As far as we can see, this undermines the support Kahan et al.

[1] Such effects go under different names in the literature; Taber and Lodge (2006) for instance detect a "sophistication effect" in their data. Our focus will be on a particular and very influential study by Kahan et al. (2017) finding an effect of precisely numeracy; thus our label. Other studies detecting such effects include McCright et al. 2016; Drummand and Fischhoff 2017; and Nurse and Grant 2020. Empirically oriented philosophers interested in (apparently) evidence resistant belief also take such effects on board; cf. for instance Levy 2019; p. 3; Levy 2022; p. 31; 126. It should be noted, however, that regarding the existence of these effects, the science is hardly settled at present – the relevant results do not seem to replicate all that easily. Among those not finding them are Tappin et al. 2021; Lind et al. 2022; Connor et al. 2020; Persson et al. 2021. See Kahan and Peters 2017 for an early reaction to claims of non-replication, and Stagnaro et al. 2023 for a recent study taking concerns about earlier replication studies into account, yet finding null results.

take the effect to provide for ICT. In Section 3, we conclude with some reflections on the wider significance our being right about the nature of Kahan's mechanism would have – both for epistemology and for debates concerning the nature of fact polarization and "partisan cognition" (Williams 2023; p. 4) in general.[2]

1 Identity Protective Cognition and Numeracy

Kahan explicates motivated reasoning as "the tendency of individuals to unconsciously conform assessment of factual information to some goal collateral to assessing its truth" (Kahan 2016a, 2).[3] The goal in question concerns the formation or maintenance of specific (kinds of) belief. In politically motivated reasoning, the truth-independent goal is identity protection (cf. ibid., 3), and it is achieved by a form of selectively "skewed" information processing: "individuals are adjusting the weight assigned evidence conditional on its identity congruence" (ibid., 10).[4] Let's call this the "IP-account" of politically motivated reasoning.

According to Kahan, identity protective reasoning drives fact polarization because certain factual claims have become symbols of membership in what he calls "identity defining affinity groups". Examples would be climate change denial and covid vaccine skepticism; these seem to have become "identity markers" in certain conservative groups in the US. Relinquishing an identity marker can carry heavy social and even economic costs. Group members thus have powerful incentives to maintain belief in such markers – even in the face of strong evidence against them.

[2] "Partisan cognition" is Williams' term for cognition influenced by party allegiances (Williams 2023; p. 4). The term is explicitly coined to be neutral regarding the motivational or non-motivational nature of such influences. Williams uses it to cover cognition about politics in general, not just the factual matters that we are interested in here, however.

[3] In the huge empirical literature on motivated reasoning there is little consensus on what motivated reasoning precisely is. A classical survey is Kunda 1990. Most agree that motivated reasoning includes, but is wider than the folk psychological notion of wishful thinking. The relevant motivation is usually understood in terms of "directional" goals, but there is no consensus on whether an affective component needs to be involved. Importantly, truth or accuracy are often counted among the directional goals. Thus, some hold that all reasoning is motivated, even reasoning solely motivated by truth or accuracy (cf. for instance Taber & Lodge 2006, p. 756.) But, like Kahan, many others require the goals pursued in motivated reasoning to precisely be *"collateral"* to truth or accuracy.

[4] Kahan is keen on distinguishing politically motivated reasoning from confirmation bias (classically surveyed in Nickerson 1998). Using Bayesian processing as a foil, Kahan construes both motivated reasoning and confirmation bias in terms of selectively adjusting your conditional probabilities ("likelihood ratios") to suit your goals. What is different are the goals: In confirmation bias, he maintains, a subject selectively adjusts conditional probabilities with the goal of generating updates consistent with prior belief. In politically motivated reasoning, by contrast, "it will look like she is deriving the likelihood ratio from her priors. However, the correlation is spurious: a 'third variable'—her motivation to form beliefs congenial to her identity—is the 'cause' of both her priors and her likelihood ratio assessment" (Kahan 2016a, 5).

The threat such *"identity-incongruent"* evidence poses to one's identity can be averted by engaging in politically motivated reasoning. To repeat, in such reasoning "individuals are adjusting the weight assigned to the evidence conditional on its identity congruence" (Kahan 2016a, 10). A threat from evidence for climate change, for instance, can be averted by (unconsciously) adjusting the evidential value of such information in a way allowing for maintaining climate denial. In general, identity-incongruent information thus can be resisted by processing it as if it had lower evidential value than it would be (epistemically) rational to assign to it.[5]

To test whether politically motivated reasoning on the IP-account explains fact polarization, Kahan has developed what he calls the "politically motivated reasoning paradigm (PMRP)" (cf. Kahan 2016a, b):

> The best test of politically motivated reasoning is whether study subjects alter the weight they assign the same piece of evidence in response to an experimental manipulation of the perceived relationship between that evidence and positions that predominate in their cultural group. This experimental setup can be called the PMRP design (Kahan 2016b; p. 2).

In experiments of this design – *"outcome switching"* – subjects' evaluation or endorsement of new information is used as outcome variable. Subjects are randomly assigned to receive one of two pieces of information. The substantive detail of the information is held constant across conditions. But its implication for subjects' political identities is varied between conditions: In one condition, the information is identity-conguent, in the other it is incongruent. The key result is that subjects'

[5] Kahan himself claims that politically motivated maintenance of an identity marker can be "perfectly rational" (Kahan 2016a, 2). The kind of rationality he has in mind, however, is of a practical nature. It is a matter of weighing risks and costs of a predominantly social and economic kind, and for those heavily invested in belonging to a certain group, the costs of giving up any of the group's identity markers clearly outweigh those of maintaining them, Kahan et al. argue:

> On matters like climate change (…) an ordinary citizen pays no price for forming a perception of fact that is contrary to the best available empirical evidence: that individual's personal beliefs and related actions – as consumer, voter, or public discussant – are too inconsequential to affect the level of risk that person or anyone else faces or the outcome of any public policy debate. However, if that person gets 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, stigmatisation within his community, and even loss of economic opportunities (Kahan et al. 2017; p. 57).

But of course, none of this will make belief resulting from politically motivated reasoning epistemically rational. It is, moreover, an interesting question whether the IP-account requires politically motivated reasoning to be practically rational. This might be so since the identity protective incentives presumably need to outweigh all opposing motivations in order to explain the resulting belief. If practical rationality is required, the account would be counter-exampled by instances of practically irrational identity protection, i.e. instances where identity markers are maintained even though identity protection is in fact outweighed by other goals. Consider covid vaccine skepticism. During the pandemic, such skepticism became an identity marker in certain conservative groups in the US. But it pretty obviously could cost your life. The fairly high rate at which people seem to have risked their lives by abstaining from vaccination rather than abandon their skepticism arguably casts some doubt on the idea that identity protection necessarily is practically rational.

evaluation of the information differs by condition, and, in particular, that this difference is correlated with their political identities or preferences.[6]

In the quite famous "skin cream/gun control" study that will be central to our concerns (Kahan et al. 2017), for instance, the information consists of a table with numerical values. What varies is what these values are described as representing. In the two control conditions, they are supposed to represent the results of a scientific study concerning the effectiveness of a skin cream in treating a certain kind of rash. What is switched is the description of what the numbers represent: people whose rash got better vs. people whose rash got worse. Participants then are asked which of the following two claims are supported by the study: People who used the skin cream were more likely to get better/get worse than those who didn't. In the experimental conditions, the very same numbers are described as results of a scientific study investigating the effects of banning the carrying of concealed handguns on violent crime. And again, outcomes are switched between conditions and participants are asked which of two claims were supported by the study: Cities that enacted a ban on carrying concealed handguns were more likely to have a decrease/increase in crime than cities without bans. The results were clear: As opposed to the control conditions, subjects' evaluation of the information differed between experimental conditions in a way correlated with their political identities.

As we have described the results so far, however, Kahan et al. do not seem to find them to be quite enough to support ICT over what they appear to consider its main rival: "the science comprehension thesis (SCT)":

> ICT, in an important sense, stands SCT on its head. Whereas SCT attributes conflicts over decision-relevant science to deficits in science comprehension, ICT sees the public's otherwise intact capacity to comprehend decision-relevant science as disabled by cultural and political conflict" (Kahan et al. 2017; p. 56).

Deciding between ICT and SCT clearly is of some practical importance; effectively counteracting fact polarization presumably takes quite different forms depending on which is the correct explanation. What is less clear is why Kahan et al. would consider both ICT and SCT to be compatible with the observed tendencies to evaluate information in ways correlated with political identity. Prima facie at least, political identity would not seem to play much of a role in the SCT framework, after all. We need to dig a little deeper here.

[6] It should be noted that designing experiments diagnostic of (politically) motivated reasoning is not an easy task. The PMRP design reflects Kahan's particular concern with, and attempt at, disentangling politically motivated reasoning from confirmation bias (see fn. 3 above).

But even the PMRP design has been compellingly argued to be importantly confounded: As subjects' political identities tend to correlate with their prior beliefs on politicised matters, the results might be due to the effects of prior belief rather than politically motivated reasoning. Cf. Tappin et al. 2020; 2021. Such prior belief might of course itself result from politically motivated reasoning, but it might also be perfectly rational. The problem is that the experiments do not allow us to tell. This kind of confound arguably afflicts experimental research on (politically) motivated reasoning in general and has been called "an empirical catch-22 at [its] heart" (Ditto et al. 2018; p. 285).

SCT attributes polarization over risks and other policy-relevant facts "to deficits in the public's capacity to comprehend scientific evidence" (Kahan et al. 2017; p. 56). According to SCT, a weak understanding of science makes ordinary citizens liable to misunderstand what scientists say as well as vulnerable to being misled about it by interested parties. The effects of limited knowledge about science are aggravated by the heuristic-driven form of information processing ("system 1" reasoning, cf. Stanovich & West 2000; Kahnemann 2003) thought to be typical of human risk perception:

> Over-reliance on System 1 heuristics is the root of myriad cognitive biases. By fixing attention on emotionally gripping instances of harm, or by inducing selective attention to evidence that confirms rather than disappoints moral predispositions, System 1 information processing induces members of the public variously to overestimate some risks and underestimate others relative to the best available evidence, the proper evaluation of which requires exercise of more deliberate and reflective 'System 2' forms of information processing (Kahan et al. 2017; p. 56).[7]

As characterized by Kahan et al., SCT would indeed seem to predict difficulties with interpreting scientific evidence. But as these would seem to be difficulties with interpreting evidence quite generally, one might still wonder whether the theory has resources to explain difficulties differentially affecting the interpretation of identity-incongruent evidence.[8]

Kahan et al. do not explicitly address this question, but we find it quite plausible to think that identity-(in)congruence does have significance in the framework of SCT, if only indirectly so. What gets challenged in Kahan et al.'s experimental conditions are identity markers, i.e. prior beliefs possessing a certain social significance. According to Kahan et al., a challenge to such beliefs affects the interpretation of the challenging evidence in the specific way characteristic of motivated reasoning. SCT, it seems to us, at least suggests alternative, non-motivational ways in which such challenges might affect the interpretation of the evidence.[9] After all, the relevant prior beliefs presumably are held quite strongly. When presented with tables of numerical values supposedly representing the results of a scientific study, subjects lacking in the ability to interpret such evidence might for instance just fall back on their strongly held prior beliefs. Moreover, as Kahan et al. point out, a lack in science comprehension is apt to make people more vulnerable to being misled into science-skeptical attitudes, thus becoming prone to disregarding scientific evidence in favor of their prior beliefs or "gut feelings". And even subjects with some

[7] For further references, see Kahan et al. 2017, p. 56.

[8] Thanks to Steven Verheyen for prompting us to reflect on precisely how SCT would explain this.

[9] As noted above (fn. 5), in outcome switching experiments the effects of motivation appear confounded by those of prior belief, a problem in fact shared by all "matched information designs" (Williams 2023; p. 8, cf. Tappin et al. 2020). Moreover, there is a quite general "problem of observational equivalence" (Williams 2023; p. 12) afflicting the interpretation of both observational and experimental findings regarding partisan cognition: both motivational and non-motivational accounts are available and appear to be explanatorily equivalent. In this wider context, SCT is one example of a (family of) non-motivational accounts of partisan cognition.

science literacy might fall prey to various biases favoring strongly held prior beliefs, biases like confirmation bias or "belief bias" (Tappin et al. 2020; p. 84). In sum, it does not seem implausible to think that both SCT and ICT are compatible with the results of the skin cream/gun control experiment as described so far.

It is here that the numeracy effect finally comes in: the skin cream/gun control study is designed to test a particular prediction distinguishing ICT from SCT. This prediction concerns the particular shape the data will be taking. According to Kahan et al., ICT predicts a numeracy effect; it predicts that "ideological polarization (…) should be most extreme among those highest in numeracy" (Kahan et al. 2017; p. 66). The idea is this:

> [M]ore numerate individuals have a cognitive ability that lower-numeracy ones do not. ICT predicts that more numerate individuals will use that ability opportunistically in a manner geared to promoting their interests in forming and persisting in identity-protective beliefs (Kahan et al. 2017; p. 75).

By contrast, no such prediction comes with SCT. If anything, SCT would predict the opposite (cf. Kahan et al. 2017, p. 65f): If SCT were correct, more numerate subjects would get things right more often, regardless of whether the numbers are identity-congruent or incongruent. This would be because more numerate subjects in general will both be higher in science-literacy and more prone to System 2 reasoning.

The results of Kahan et al.'s study bear out ICT's prediction; his data do show a numeracy effect.[10] As noted above (fn. 1), the effect does not seem to replicate very well. We'll get back to that towards the end of this paper. But for now, we shall not be concerned with the numeracy effect's existence. What we would like to understand is how this effect is supposed to work. While the more numerate of course have cognitive abilities that the less numerate lack, we do not find it obvious that higher numeracy would generate "better" identity protection. What kind of "mechanism" is supposed to produce this effect?

We shall continue to focus on Kahan et al. (2017). In the next section, we shall carefully reconstruct their own interpretation of the skin cream/gun control experiment and its results. We are going to do this in order to gain a precise understanding of what we shall call "*Kahan's mechanism*", i.e. the mechanism that according to Kahan et al. explains the numeracy effect in their data. No doubt many other interesting questions can be raised, for instance about the interpretation of the skin cream/gun control study and its results. While we shall touch upon the significance of such questions towards the end of the paper, this paper is not about the correct interpretation of this study. Our main goal is modest: it is to understand Kahan's mechanism. As already indicated, we think that this investigation leads to surprising and quite intriguing results.

[10] Even though the effect appears to wear off for the extremely numerate.

2 Kahan's Numeracy Mechanism

An intuitive take on numeracy or cognitive sophistication effects on motivated reasoning goes something like this: Those with higher reasoning capacities are able to selectively bring these capacities to bear on evidence for or against beliefs they cherish. For instance, they are better at finding fault with arguments or evidence against their views. They are also better at finding further arguments or evidence for their views. Their higher reasoning capacities thus will make them "better" at both undervaluing evidence against, and at overvaluing evidence for, their cherished beliefs.

This intuitive take is not uncommon among those researching these matters. Taber and Lodge, for instance, predict (and find) a "sophistication effect" described as follows:

> [T]he politically knowledgeable, because they possess greater ammunition with which to counterargue incongruent facts, figures, and arguments, will be more susceptible to motivated bias than will unsophisticates (Taber & Lodge 2006, 757).

And Kahan himself comments on reasoning under identity threat:

> Under these conditions, it is a perfectly rational thing for one to attend to information in a manner that promotes beliefs that express one's identity correctly, regardless whether such beliefs are factually correct (…). And if one is really good at conscious, effortful information processing, then it pays to apply that reasoning proficiency to give information exactly this effect (Kahan 2016b; p. 4).

There is, however, something quite puzzling about this kind of idea. It can hardly be the case that more numerate individuals are better at consciously and deliberately engaging in lowering (or raising) the evidential value of a piece of evidence in order to maintain some cherished belief. Rather, such maintenance would seem to require that at least some of the reasoning remains hidden from the subject's own view – quite irrespective of their numeracy levels.

In order to understand how this is exactly supposed to work, we are now going to take a careful look both at Kahan et al.'s skin cream/gun control study (Kahan et al. 2017) and at the details of the account of the numeracy effect provided in that paper. The design of the skin cream/gun control study study is quite sophisticated. The table shown to participants in one of the experimental conditions looks as follows (Fig. 1).

In the other experimental condition, the labels on the columns are switched (Fig. 2).

The numbers in this table are *"difficult"*, i.e. they are such that the correct answer isn't obvious (to most of us). To figure it out, we need to calculate the percentage of cities in each category that experienced an increase/decrease in crime, which requires summing up the numbers in each row first. In Fig. 1, for instance, there are 298 cities that banned concealed handguns. Of these, 75 (roughly 25%) experienced a decrease in crime. And there were 128 cities that did not have a

	Result: Increase in crime	Result: Decrease in crime
Cities that did ban carrying concealed handguns in public	223	75
Cities that did not ban carrying concealed handguns in public	107	21

What result does the study support?

Fig. 1 Table from Kahan et. al. 2017

	Result: Increase in crime	Result: Decrease in crime
Cities that did ban carrying concealed handguns in public	223	75
Cities that did not ban carrying concealed handguns in public	107	21

What result does the study support?

Fig. 2 Label switching

ban, of which 21 (roughly 16%) experienced such a decrease. In Fig. 1, the correct answer thus is that the study supports the conclusion that cities that enacted a ban on carrying concealed handguns were more likely to have a decrease in crime than cities without bans.

But the numbers are not only difficult, they are also designed to be *"misleading"*: They are such that either of the two heuristics that people tend to use when answering such questions will lead to the wrong answer. That is, if we skip the laborious system 2 reasoning just illustrated and go for what system 1 suggests, we will get it wrong. With respect to tables like these, what we shall call *"the table heuristics"* consist in either just comparing the numbers in the upper row, or in comparing the number in the upper left corner with that in the lower left corner (Fig. 3; cf. Kahan et al. 2017, 61f).

Because of the table heuristics, these numbers "look" misleading to most people: As illustrated in Fig. 3, either one of them leads to the wrong conclusion that cities that banned handguns were more likely to see an increase in crime. In sum, arriving at the correct answer requires actually doing (or at least approximating) the calculation. The study thus is designed to test for a numeracy effect by using numbers that are both difficult and misleading.

Predictably, people on average don't do extremely well in this task even in neutral conditions. More numerate subjects – also predictably – perform better in the

Fig. 3 Heuristics

neutral conditions. And they also perform better in the experimental condition congruent with their political identity. But in Kahan et al.'s study, they perform just as badly as everyone else in identity-incongruent conditions.

Kahan et al.'s explanation of this numeracy effect turns on both the misleadingness and the difficulty of the numbers. As far as we can tell, it is supposed to work as follows (cf. Kahan et al. 2017; p. 75): The initial use of a heuristic will make a condition that in fact is identity-congruent look identity-incongruent. This puts the subject under identity threat. But since the numbers are difficult, it seems possible to the subject that they in fact are not incongruent. They will therefore be motivated to check this by doing a proper calculation, and those higher in numeracy (more often) will get it right. The identity-incongruent condition, by contrast, will, by the heuristic, look congruent. Lacking any incentive to calculate, most subjects – whatever their numeracy score – will be content with the answer delivered by the heuristic. All of these subjects will get it wrong.

Kahan et al.'s explanation of the numeracy effect is quite intriguing, but it does raise the question: Where is the motivated reasoning in this mechanism? And as far as we can tell, there isn't any.

To begin with, Kahan's mechanism is rather unlike the intuitive idea we started this section with. According to Kahan et al., what the subjects with higher numeracy scores do when they feel under identity threat is not some sophisticated kind of counterarguing. They just calculate. The motivation to do so being the same for all subjects, the numeracy effect here really is nothing but the effect higher numeracy usually has when calculating: getting it right more often.

Moreover, when under identity threat, the subjects higher in numeracy calculate in precisely the same way they do in the neutral conditions. The motivation deriving from the perceived identity threat does not interfere with or "skew" their information processing. Nor does the motivation result in the signature effect of motivated reasoning as understood by Kahan: the selective (and epistemically irrational) adjusting of the evidential value of the information processed. After all, the numeracy effect comes about precisely because the more numerate subjects when feeling under threat get things right. As far as we can see, there is no motivated reasoning in this. What there is is motivation to reason. But that is a very different thing.

If what generates the numeracy effect indeed is an interaction between motivation *to* reason and numeracy, it is hard to see how the effect would support ICT over rival explanations such as SCT. This is just because it would not support ICT to begin with. To do so, the effect would have to be due to identity protective motivated reasoning, and so far, it does not look as if it were. If the effect is due to Kahan's mechanism, what identity threat generates in those higher in numeracy is not motivated reasoning at all, but quite the opposite: correct calculations.[11]

There is, however, an additional, somewhat overlooked aspect of the mechanism. So far, we have seen how it is supposed to explain why subjects calculate in incongruent looking conditions: they feel under identity threat. But this is only half the story, so to speak: we also need an explanation for why those higher in numeracy do not calculate in the experimental conditions where they don't feel under identity threat. Kahan et al. themselves appear to think that no further explanation is required here: Since no identity threat is perceived, there is no motivation to calculate.

But this won't do. Those higher in numeracy are assumed to be inherently more motivated to do the relevant calculations. As Kahan et al. explain, "the numeracy scale measures a disposition to subject intuition to critical interrogation in light of all available information, and thus to avoid mistakes characteristic of overreliance on heuristic, System 1 information processing" (Kahan et al. 2017; p. 64, with reference to Liberali et al. 2012). And as we saw, many of those higher in numeracy do the calculations in the neutral conditions of the skin cream/gun control study. That's why they get it right more often in those conditions, after all. So, more numerate subjects are typically motivated to calculate not only when under identity threat, but also in conditions without any such threat. Neutral conditions clearly fall in the latter category, but so do experimental conditions in which no identity threat is perceived – prima facie at least. That more numerate subjects do not calculate in experimental conditions where no identity threat is perceived thus requires explanation.

At this point, it might even seem as if the more numerate subjects after all did have a particular ability beyond their numeracy: They appear to be able to "switch off" their numeracy in conditions where using it would not serve their identity protective needs. This is how for instance Neil Levy interprets Kahan et al.:

> Kahan suggests that the greater polarization seen among more capable and informed participants is due to their greater capacity. This capacity gives them an ability less capable participants don't possess: to clearly recognize how threatening the correct response is to their world-view, or identity. They are therefore motivated to selectively inhibit Type 2 cognition (Levy 2022; p. 31).

The problem is that this gets the experiment wrong, at least according to Kahan et al.'s own take on it: in the skin cream/gun control study, the motivation provided by (perceived) identity threat is not motivation to inhibit system 2, on the contrary, it is motivation to switch it on. While we indeed need an explanation for why high numeracy subjects

[11] This is not to say that the effect, as explained by Kahan's mechanism, supports SCT instead. If SCT indeed predicts the absence of a numeracy effect, the presence of such an effect would be evidence against SCT, no matter what explains it.

"inhibit Type 2 cognition" in congruent looking conditions, recognized identity threat can't be it. The conditions are designed precisely to keep the threat hidden. To recognize it, the calculation has to already have been carried out. In other words, to recognize the real, but hidden threat, the subjects would need to do precisely the opposite of what they appear to be doing: use, instead of inhibit, system 2.[12]

As Kahan et al. themselves appear to think that the absence of perceived threat suffices as an explanation here, we can only speculate as to what a better one might be. Presumably, it would have to do with political charge. In the experimental conditions, the topic is one on which the subjects presumably have a desire not to receive identity-incongruent information. They might thus refrain from calculating out of apprehension of receiving counterevidence.

If this were the explanation, it would be quite ironic. For again, the motivation provided here is the same for all subjects. The "ability" the less numerate lack would then be the ability to inhibit higher numeracy skills when apprehending counterevidence. But putting it like this is misleading. If apprehension of counterevidence provides motivation to not calculate, all subjects have that motivation. And all subjects have the ability to not even try to calculate.[13]

More importantly, if calculation is refrained from because of apprehension of counterevidence, the question returns: Where is the motivated reasoning in this mechanism? As far as we can tell, there isn't any. The motivation provided by apprehension of counterevidence does not interfere with, or "skew", any reasoning process. Rather, what takes place is the quite normal initial employment of a heuristics. What does not take place is further reasoning. There is motivation, but it does not result in motivated reasoning.[14] It is motivation to refrain from reasoning, and again, that is a very different thing.[15] So, no matter how we turn it, we cannot find any motivated reasoning in Kahan's mechanism.

[12] A similar mistake appears to be made by Connor et al. when they interpret the skin cream study results as follows:

[H]igh-numeracy liberals tended to successfully reason about accuracy only when the evidence suggested that gun control is effective, whereas high-numeracy conservatives tended to successfully reason about accuracy only when the evidence suggested that gun control is not effective (Connor et al. 2020; p. 25).

While it is true, of course, that subjects higher in numeracy succeeded when – objectively speaking – the evidence suggested the identity-congruent conclusion, Connor et al. – just like Levy – miss the fact that the evidence did not suggest this to the study participants. What the evidence is designed to suggest to the subjects is the opposite of what it actually, objectively supports. Since they miss this crucial feature of the study design, commentators like Levy and Connor et al. also miss the fact that Kahan et al. do not provide a satisfactory explanation for why subjects high in numeracy refrain from calculating when not feeling under identity threat.

[13] Maybe subjects higher in numeracy typically are more motivated to calculate. Their decision to not calculate would then require greater willpower, so to speak. But that, too, wasn't the kind of greater ability supposed to produce the numeracy effect.

[14] In the skin cream/gun control study, the heuristics will lead to incorrect answers. That is, the subjects will arrive at the same conclusions that would be delivered by undervaluing the counterevidence the numbers actually provide. But these conclusions will not be the result of any actual (system 2) information processing involving undervalued evidence; they will be the result of resting content with what a perfectly normal (system 1) heuristics delivers.

[15] If not calculating indeed is the result of acting on motivation to refrain from calculating, the process is more like what media scientists call "selective exposure", i.e. the active choice to avoid obtaining information from certain sources.

As we said above, if the effect is generated by an interaction between motivation to reason and numeracy, it is hard to see how it would support ICT over rival non-motivational explanations of fact polarization (such as SCT). This verdict does not change if the mechanism not only involves motivation to reason, but also motivation to refrain from reasoning. Either way, if it is Kahan's mechanism that generates it, the effect is not due to motivated reasoning at all, and, consequently, its existence does not provide any (additional) support for ICT.[16]

3 Conclusion

According to Kahan, identity-protective reasoning is the main driver of the political fact polarization increasingly affecting political life (at least in the Western world). One crucial piece of evidence for the identity-protective cognition thesis (ICT) consists of experimental data supporting the claim that ideological polarization is more extreme amongst more numerate individuals. In this paper, we have taken a close look at how precisely this numeracy effect is supposed to come about. Identity-protective reasoning supposedly being a (political) kind of motivated reasoning, we have used Kahan's own notion of motivated reasoning to carefully reconstruct the mechanism that according to him and his colleagues produces the effect. Surprisingly, it has turned out to involve plenty of motivation to reason (and, possibly, to refrain from reasoning), but no motivated reasoning.

In these concluding remarks, we would like to add some reflections on the wider significance of this result, both from the perspective of epistemology and from that of the debate concerning motivational and non-motivational explanations of fact polarization and partisan cognition more generally. Concerning fact polarization, Kahan himself clearly considers the results obtained by means of the PMPR paradigm, and the numeracy effect in particular, as a bit of a silver bullet disentangling the motivational from the non-motivational. But as we saw, if it is Kahan's mechanism that generates it, the numeracy effect is not due to motivated reasoning at all. Consequently, its existence provides no (additional) support for either ICT or the motivated nature of partisan cognition in general.

We'll return to this debate in a moment. But first, we would like to take a step back and consider some questions concerning rationality and reasoning, questions we have so far only alluded to. As we observed earlier, identity-protective reasoning might be rational in some practical sense, but it certainly is irrational in the epistemic sense. This does not mean, however, that non-motivational accounts of fact polarization automatically construe the partisan cognition involved as epistemically rational, far from it. The distinction between these two families of accounts rather turns on the question whether the (typically unconscious) reasoning (or belief formation process) in question is purely "cognitive" or not. In motivated reasoning it is not; here, the process is "skewed" – interfered with – by motivational states

[16] As already noted (fn. 10), this does not mean that the effect would support SCT over ICT, however.

(desires), and thereby irrational. According to non-motivational accounts, by contrast, the relevant reasoning involves cognitive states only. Nevertheless, it might be subject to various biases or use heuristics making it less than rational.[17] Whether motivational or non-motivational, accounts involving irrationality typically are such that the irrationality results from some kind of systematic tendency or mechanism.

Interestingly, Kahan's mechanism is quite different. In a sense, it cuts across the motivational/non-motivational distinction: It does not involve motivational interference with belief formation, but it does involve motivational states. The motivation, however, is just your quite normal, practical motivation: it is motivation to act in a certain way (or to refrain from so acting). Moreover, the reasoning subjects are motivated to perform will be (fairly complicated) system 2 reasoning, which tends to be conscious and "explicit". The mechanism generates a numeracy effect simply because less numerate subjects are more error-prone when engaging in such reasoning. Calculating mistakes do result in epistemically irrational beliefs, but this is not the kind of systematic tendency or mechanism supposed to do the explanatory work in "irrationality accounts" of partisan cognition.

Now, in what we shall call "calculating cases", i.e. cases in which subjects have motivation to calculate, there is a clear sense in which doing so is practically rational (as long as they don't have any stronger reasons against calculating). And analogously for subjects in "refraining cases", i.e. subjects having motivation to refrain from calculating. Keeping the distinction between practical and epistemic irrationality strict, one might then worry that while Kahan's mechanism can give rise to epistemic irrationality, it can do so only in calculating cases (where it moreover is restricted to calculating mistakes). Intuitively, this might not seem quite right; there seems to be more irrationality than just that. Let's call this the "more irrationality intuition".

There is a lively debate in current epistemology that would seem relevant to this intuition.[18] It concerns the epistemic relevance of evidence a subject does not have, but could easily have obtained. A growing number of epistemologists argue that such evidence can determine whether beliefs it is relevant for are epistemically rational (or justified), and that therefore at least certain "zetetic norms" are epistemic despite their practical character (cf. Goldberg 2017; Ichikawa 2022; Flores & Woodward 2023, Simion 2023, 2024). Zetetic norms are norms of inquiry, in this case norms for gathering evidence. Consider a piece of evidence *e* that a subject *S* does not have, but could easily obtain. Roughly, the basic idea is that under certain circumstances, *S* "should" have *e*. Under such circumstances, refraining from getting *e* is epistemically irrational and maintaining or forming a belief that is (or would be) defeated by *e* is unjustified.[19] On such a picture, we might be able to account for the more irrationality intuition: Our refraining cases might be such that the subjects

[17] Here, we largely follow the survey of both motivational and non-motivational accounts of partisan cognition provided in Williams 2023, p. 4 ff.

[18] Thanks to an anonymous referee for prompting us to connect our discussion with this debate.

[19] Simion (2023, 2024) builds an interesting case not only for the epistemic significance of evidence subjects "should" have had, but also for a particular, factive account of the nature of evidence precisely on cases of "knowledge or evidence resistance". These are cases where subjects' not taking in easily available evidence negatively affects the epistemic status of (a relevant subset of) their beliefs.

"should" have at least tried to obtain the evidence, i.e. to calculate. Refraining then would itself be epistemically irrational and the beliefs formed, or held on to, despite the availability of evidence against them unjustified.

This is certainly interesting, but we can't explore it much further here. We would like to note, however, that even though calculating mistakes are the only epistemic irrationalities Kahan's mechanism can bring about, there is an alternative directly relevant source of irrationality in the subjects' responses to the numbers: the table heuristics. After all, it is their outputs that make the mechanism kick in. And in refraining cases, the mechanism allows subjects to stick with their heuristics-generated beliefs. Whether heuristics-generated beliefs in general are epistemically irrational is a matter of debate, but the particular heuristics relevant here – the table heuristics – might just be sufficiently unreliable for such a verdict. In that case, we might not need zetetic norms, or the effects of evidence not actually had by the subjects, to account for the more irrationality intuition. The relevant beliefs would be irrational already in virtue of being heuristics-generated, irrespective of any unobtained evidence "out there".

Thus far, we have considered Kahan's mechanism not only on the assumption that there *is* a numeracy effect, but also that it is generated by Kahan's mechanism. To round things off we would like to comment on the question whether Kahan et al. are right about what generates the numeracy effect – if there is one. The short answer is: We don't know. Even so, some things can be said: It does, for instance, seem fair to say that Kahan et al. provide a reasonably plausible explanation of such an effect, an explanation, moreover, that cannot be ruled out on the basis of experiments such as the skin cream/gun control experiment alone. That by no means excludes the possibility of other plausible explanations.[20] Nor can it be ruled out that the correct explanation might yet involve motivated reasoning. But that none of these possibilities is ruled out by them suffices to undermine the claim that the results of such experiments provide additional support for ICT. So, for undermining the support a numeracy effect could provide for ICT (and, thereby, the motivated nature of partisan cognition) we do not need to assume that Kahan's mechanism would be the correct explanation – all we need is that this explanation is a "live option", an option that has not been ruled out.

By itself, the existence of the numeracy effect – far from being a silver bullet – thus is quite powerless to take us out of the impasse many think the debate between motivational and non-motivational accounts of fact polarization, and partisan cognition more generally, has reached (cf. Williams 2023; p. 89).[21] In a sense, the question whether the

[20] It does not seem implausible, for instance, that more numerate subjects also are better informed subjects. Better informed subjects will (take themselves to) have strong justification for their prior beliefs, especially strongly held ones, and might for instance take evidence challenging these beliefs to in fact cast doubt on the reliability of the evidence's source. To the extent that their prior beliefs actually are justified such subjects' responses might be perfectly rational. But even if the justification is spurious, the resulting reasoning clearly would not be motivated reasoning and yet give rise to a numeracy effect (cf. Tappin et al. 2021; Stagnaro et al. 2023; p. 2).

[21] Depending on how one assesses the overall balance of available evidence, and especially the weight of that provided through Kahan's paradigm, one might think that "neutralizing" the numeracy effect tips the balance towards non-motivational accounts. This is nothing we can take a stand on based on what we have said in this paper.

numeracy effect actually exists therefore becomes moot. Even if it replicated, which at the time of writing appears rather doubtful (see above, fn. 1), some other means of breaking the impasse would be needed.[22] Thinking about how to best go forward is beyond the scope of this paper, however. For now, we conclude that on careful examination, Kahan's mechanism turns out to involve motivation to reason (and, possibly, to refrain from it), but no motivated reasoning, and that this undermines the support he takes the numeracy effect to provide for the identity protective cognition hypothesis.[23]

Acknowledgements For valuable comments, the authors would like to thank audiences in Stockholm, Glasgow, Lund, and Aberdeen, and especially Åsa Wikforss, Peter Pagin, Roderik Rekker, Mona Simion, Chris Kelp, Adam Carter, Luca Moretti, and Jesper Kallestrup. They would also like to thank their editor, Steven Verheyen, and two anonymous reviewers for very helpful and instructive comments.

Funding Open access funding provided by Stockholm University. This work was supported by the interdisciplinary research program "Knowledge Resistance: Causes, Consequences, and Cures", funded by Riksbankens Jubileumsfond (RJ M18-0310:1). The paper was revised while Kathrin Glüer-Pagin was a fellow at SCAS, the Swedish Collegium for Advanced Study.

Declarations

Conflict of Interests The authors have no relevant financial or non-financial interests to disclose.

References

Connor, P., et al. 2020. Motivated numeracy and active reasoning in a Western European Sample. *Behavioural Public Policy 2024* 8 (1): 24–46.

Ditto, Peter H., et al. 2018. At Least Bias Is Bipartisan: A Meta-Analytic Comparison of Partisan Bias in Liberals and Conservatives. *Perspectives on Psychological Science* 14 (2019): 273–291.

[22] This might be yet another experimental design, of course, but as noted before (fn. 6), it is hard to come up with experiments truly diagnostic of politically motivated reasoning. Williams 2023 therefore argues that in order to make progress here "we need a more developed theory of the connections between group attachments and motivated reasoning" (2023,1). He makes an interesting proposal as to how to do this, but discussion of "the coalitional press secretary theory of partisan cognition" is beyond the scope of this paper.

[23] Yet another question is whether Kahan's mechanism has any significance for these debates quite independently of numeracy effects. One might suspect so. Motivation to reason (and to abstain from reasoning) might be (part of) the explanation of tendencies to misinterpret identity-incongruent evidence quite in general, or at least when it takes certain numerical shapes. After all, in Kahan's experiment all subjects show tendencies to interpret the numbers correctly when they in fact are identity-congruent, and incorrectly otherwise. And this finding does replicate (cf. especially Stagnaro et al. 2023 where the tendency is shown across a range of six different politically polarized issues, but diminishes with higher numeracy). Again, this deserves further exploration, but cannot be pursued here.

Drummand, C., and B. Fischhoff. 2017. Individuals with greater science literacy and education have more polarized beliefs on controversial science topics. *Proceedings of the National Academy of Sciences* 114: 9587–9592.

Flores, Carolina, and E. Woodard. 2023. Epistemic norms on evidence-gathering. *Philosophical Studies* 180: 2547–2571.

Goldberg, Sanford C. 2017. Should have known. *Synthese* 194: 2863–2894.

Ichikawa, Jonathan J. 2022. You ought to have known: positive epistemic norms in a knowledge-first framework. *Synthese* 200: 1–23.

Kahan, Dan M. 2013. Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making* 8: 407–424.

Kahan, Dan M. 2015. The expressive rationality of inaccurate perceptions. *Behavioral & Brain Sciences* 40: 26–28.

Kahan, Dan 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*, ed. Robert Scott, Marlis Buchmann, and Stephen Kosslyn, 1–6. John Wiley & Sons.

Kahan, Dan M. 2016b. The politically motivated reasoning paradigm, part 2: unanswered questions. In *Emerging Trends in the Social and Behavioral Sciences*, ed. R. Scott, M. Buchmann, and S. Kosslyn, 1–15. John Wiley & Sons.

Kahan, Dan M., and Ellen Peters. 2017. Rumors of the 'nonreplication' of the 'motivated numeracy effect' are greatly exaggerated'. *Yale Law and Economics Research Paper* 584. https://ssrn.com/abstract=3026941 https://doi.org/10.2139/ssrn.3026941.

Kahan, Dan M., Ellen Peters, Erica Cantrell Dawson, et al. 2017. Motivated numeracy and enlightened self-government. *Behavioural Public Policy* 1: 54–86.

Kahan, Dan M., Ellen Peters, and Maggie Wittlin. 2012. The polarizing impact of science liter- acy and numeracy on perceived climate change risks. *Nature Climate Change* 2: 732–735.

Kahnemann, D. 2003. Maps of bounded rationality: psychology for behavioral economics. *American Economic Review* 93: 1449–1475.

Klein, Ezra. 2020. *Why we're polarized*. New York: Simon & Shuster.

Kunda, Ziva. 1990. The case for motivated reasoning. *Psychological Bulletin* 108: 480–498.

Levy, Neil. 2019. Due deference to denialism: explaining ordinary people's rejection of established scientific findings. *Synthese* 196: 313–327.

Levy, Neil. 2022. *Bad beliefs. Why they happen to good people*. Oxford: Oxford University Press.

Liberali, J.M., et al. 2012. Individual Differences in Numeracy and Cognitive Reflection, with implications for biases and fallacies in Probability Judgment. *Journal of Behavioral Decision Making* 25: 361–381.

Lind, Thérèse., et al. 2022. Motivated reasoning when assessing the effects of refugee intake. *Behavioural Public Policy* 6: 213–236.

McCright, A.M., et al. 2016. Ideology, capitalism, and climate: explaining public views about climate change in the United States. *Energy Research and Social Science* 21: 180–189.

Nickerson, Raymond S. 1998. Confirmation Bias: a ubiquitous phenomenon in many guises. *Review of General Psychology* 2: 175–220.

Nurse, M.S., and W.J. Grant. 2020. I'll see it when I believe it: motivated numeracy in perceptions of climate change risk. *Environmental Communication* 14: 184–201.

Persson, Emil, et al. 2021. A preregistered replication of motivated numeracy. *Cognition* 214: 104768.

Simion, Mona. 2023. Resistance to evidence and the duty to believe. *Philosophy and Phenomenological Research* 108: 203–216.

Simion, Mona. 2024. *Resistance to evidence*. Cambridge: Cambridge University Press.

Stagnaro, M.N., et al. 2023. No association between numerical ability and politically motivated reasoning in a large US probability sample. *Proceedings of the National Academy of Sciences* 120: e2301491120. https://doi.org/10.1073/pnas.2301491120.

Stanovich, K.E., and R.F. West. 2000. Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences* 23: 645–665.

Taber, C.S., and M. Lodge. 2006. Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science* 50: 755–769.

Tappin, Ben M., Gordon Pennycook, and David G. Rand. 2020. 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.

Tappin, Ben M., Gordon Pennycook, and David G. Rand. 2021. Rethinking the Link between Cognitive Sophistication and politically motivated reasoning. *Journal of Experimental Psychology: General* 150: 1095–1114.

Williams, D. 2023. The case of partisan motivated reasoning. *Synthese* 202: 89. https://doi.org/10.1007/s11229-023-04223-1.