Synthese
https://doi.org/10.1007/s11229-019-02515-z

Evidentialism and belief polarization

Emily C. McWilliams[1]

Received: 15 July 2017 / Accepted: 9 December 2019

Abstract
Belief polarization occurs when subjects who disagree about some matter of fact are exposed to a mixed body of evidence that bears on that dispute. While we might expect mutual exposure to common evidence to mitigate disagreement, since the evidence available to subjects comes to consist increasingly of items they have in common, this is not what happens. The subjects' initial disagreement becomes more pronounced because each person increases confidence in her antecedent belief. Kelly (2008a) aims to identify the mechanisms that underlie this phenomenon and assess whether these processes undermine the justification of polarized beliefs. He concludes that given evidentialism, justification is not undermined by the polarizing mechanisms. I take on board Kelly's description of the polarizing mechanisms, but challenge his conclusion. I argue that on plausible versions of evidentialism, the beliefs that result from these routes to polarization are not justified.

Keywords Evidentialism · Rationality · Justification · Belief polarization

1 Introduction

This paper concerns a type of *belief polarization* that occurs when subjects who disagree about some non-straightforward matter of fact are exposed to a mixed body of evidence that bears on that dispute. We might expect that mutual exposure to the same evidence would mitigate disagreement, since the evidence available to subjects comes to consist increasingly of items they have in common. But numerous studies cast doubt on this expectation. In many (though not all) studies where subjects are exposed to mixed evidence, the opposite happens: disagreement becomes even *more* pronounced, as each person increases confidence in her antecedent belief. Subjects' beliefs *polarize* with respect to one another.

In a well-known paper entitled "Disagreement, Dogmatism, and Belief Polarization", Tom Kelly aims to identify the mechanisms that underlie polarization in

✉ Emily C. McWilliams
emily.mcwilliams@gmail.com

[1] Duke Kunshan University, No. 8 Duke Avenue, Kunshan 215316, Jiangsu, China

Published online: 14 December 2019

certain cases and to assess the normative, epistemic issues that arise from it. In particular, he is interested in whether the processes that underlie polarization in these particular cases undermine the justification of resulting polarized beliefs. He assumes that subjects' beliefs *start out* justified and asks whether these routes to polarization leave the beliefs in good standing. He concludes that given evidentialism,[1] the justification of the subjects' beliefs is *not* undermined by their being the result of the mechanisms that underlie polarization.

I will take on board Kelly's description of the polarizing mechanisms but challenge his conclusion. Kelly's evidentialism is underspecified. I argue that evidentialism can be construed in ways that are more and less plausible with respect to what counts as a subject's evidence, and that on more plausible versions, the polarized beliefs Kelly describes are unjustified.

Why does it matter whether evidentialism says that polarized beliefs are justified? Ultimately, I argue that the question is significant since the version of evidentialism that Kelly would need to count these beliefs justified is out of sync with how the term *justification* should be used in both ordinary and theoretical contexts. This is a significant conclusion to draw about one of the main contenders for the correct theory of epistemic justification. Evidentialists will want to avoid this result.

2 Kelly on the psychological phenomena that underwrite polarization

In identifying the psychological phenomena that underwrite polarization, Kelly draws specifically on an empirical study done by Lord et al. (1979). They recruited subjects who disagreed about a complex empirical question. In this case, the question was whether capital punishment tends to have a deterrent effect on the commission of murder. Half of the subjects believed *p*: "Capital punishment does have a deterrent effect on the commission of murder", and the other half believed *not-p.* I'll call these propositions *Deterrent* and *Not-Deterrent*, respectively.

During the experiment, all subjects were shown the detailed results of two different studies that bore on the disputed question: one study offered support for *Deterrent*, and the other supported *Not-Deterrent*. Subjects were also shown a list of criticisms of each study, and replies to those criticisms. After they had time to reflect on this information, subjects on both sides of the issue reported that they had become more confident in their antecedent views about *p*: subjects who antecedently believed *Deterrent* now felt more confident that it was true, and subjects who believed *Not-Deterrent* felt more confident in that.

[1] Kelly does not *explicitly* identify himself as a strict evidentialist in his paper. He says only that *in paradigmatic cases*, how confident it is reasonable to be in one's belief in some proposition is a matter of how well-supported that proposition is by one's evidence (2008a, p. 623). Nonetheless, his argument seems to rely on this being such a case. He argues that polarized beliefs are rational by appealing to the idea that what it is reasonable to believe is a function of one's evidence (Ibid., p. 628). So, for our purposes, I will take it that evidentialism is the framework in which Kelly assesses what it is reasonable to believe. Whether he is an evidentialist beyond this will not mater for the issue I am debating.

Kelly identifies an empirical phenomenon that underwrites and explains the polarization of subjects' beliefs. I will call it *uneven scrutiny*. It is a specific way in which the more general phenomenon known in social psychology as *biased assimilation* can occur. *Uneven scrutiny* refers to the way that subjects processed the information in the two studies they were shown. Namely, they *scrutinized* the study that disagreed with their view (I'll call this the *uncongenial study*). They used their cognitive resources to search for flaws that might discredit the study's conclusion: problems with its methodology, variables that were not adequately controlled for, etc. Meanwhile, subjects took the congenial study's results on board as further evidence for their view, *without* scrutinizing it.[2]

Why did this uneven scrutiny lead to belief polarization? The experiment was set up so that *both* of the studies subjects were shown contained flaws that would discredit their authors' conclusions. But because they scrutinized unevenly, the subjects only found the flaws in the uncongenial studies. And because the flaws discredited the studies' conclusions, the subjects took them to be defeaters of the support that the uncongenial study would otherwise lend to the uncongenial view.

To illustrate how this setup results in polarization, consider a subject—call her *Elena*—who comes into the study believing *Deterrent*. She is shown a congenial study and an uncongenial one. Since the congenial study agrees with Elena's antecedent belief, she does not scrutinize it, but takes the findings at face value, and increases her confidence in *Deterrent*. Because the uncongenial study disagrees with her, she scrutinizes it, and discovers the fatal flaws in the study's procedure and methodology. She takes these flaws to defeat the evidence that the study would otherwise lend to *Not-Deterrent*; so, her credence in *Not-Deterrent* does not change.[3]

The finding of the Lord et al. study is robust. Indeed, it was later replicated by Houston and Fazio (1989), who found that this type of processing was particularly pronounced among subjects whose attitudes towards the death penalty were 'cognitively accessible' to them.[4]

[2] Note that by engaging in uneven scrutiny, subjects are *not* acting dogmatically in the sense that is relevant to Saul Kripke's well-known "dogmatism paradox". That is, that are not dismissing apparent counterevidence; rather, they are paying *more* attention to it. Kelly is careful to point this out in his paper, as he believes that updating in the manner of the Kripkean dogmatist is unreasonable because it violates the commutativity of evidence principle. See Kelly (2008a, pp. 3–8).

[3] This effect—of prior attitudes leading to uneven scrutiny, and polarization—seems to be robust. It has been documented more recently by Taber and Lodge (2006) in their work on political beliefs.

[4] Other studies that have found evidence of biased assimilation of mixed evidence include: Hastrof and Cantril (1954), Miller et al. (1993, Experiment 3), Kunda (1987), Chen et al. (1992), Koehler (1993), Kluegel and Smith (1986), Murno and Ditto (1997), Batson (1975) and Liberman and Chaiken (1992). There has been controversy about whether belief polarization was found in the Miller et al. (1993) and Murno and Ditto (1997) studies. In both cases, polarization was not found when beliefs were measured directly, but only when they were self-reported. If the concern is whether self-reported attitude change affects behavior, then it is worth noting that speech acts of reported attitude change are themselves a kind of behavior. There may be an impact on how we jointly deliberate and interact with one another in conversations around these issues, which is significant in itself for epistemic reasons.

3 Kelly's normative assessment

Kelly aims to evaluate the justificatory status of the beliefs that result from this route to polarization. His view about what determines a belief's justification is a standard kind of *evidentialism*, which says that whether a belief that *p* is justified is (only) a matter of whether *p* is sufficiently well-supported by the subject S's evidence.[5] A belief is justified *iff* it is supported by S's evidence.[6] Importantly, this means that how and why S gets her evidence is not relevant to the justificatory status of her beliefs. So, it does not matter that S gets her evidence from the process of uneven scrutiny I described. For Kelly, S's polarized beliefs are justified because, he thinks, those beliefs are well-supported by S's evidence.

More specifically, Kelly reasons as follows. Consider Elena again, who starts out believing *Deterrent*. When she encounters the uncongenial study, she scrutinizes it, and gains a defeater[7] for the evidence that it would otherwise give her for *Not-Deterrent*. Therefore, she does not get new evidence in favor of *Not-Deterrent*, and her disbelief in *Not-Deterrent* does not and should not change. When she encounters the congenial study, she does not scrutinize it. She remains unaware of its flaws, so her evidential perspective does not include a defeater for the evidence it seems to provide for *Deterrent*. At the end of the experiment, Elena has new information in support of *Deterrent*, and no new information in support of *Not-Deterrent*. Therefore, proportioning her belief to her evidence requires her to increase her confidence in *Deterrent*. When she does, her belief remains justified.

To get clear on what Kelly counts as part of her evidence, and how it serves his argument that her polarized belief is well-supported by her total evidence, I reconstruct the argument in more detail. In Sect. 4, I make a case that it covers over important details having to do with Kelly's conception of Elena's evidence. On Kelly's analysis of the case, Elena gets the following new evidence during the course of the experiment:

[5] See, for instance, Kelly (2003). The fundamentals of the view are taken over from Conee and Feldman (2004a, b). Kelly distinguishes two different senses of 'evidence': a *broad* and a *narrow* sense. The use of *evidence* in this formulation of evidentialism refers to evidence in the broad sense. I will discuss this distinction in Sect. 4.3.2.

[6] Strictly speaking, matters may be slightly more complicated, since to secure *doxastic* (as opposed to *propositional*) justification, the evidentialist might also say that it is necessary that the belief be appropriately based on the relevant evidence. Kelly does not say whether he is talking about doxastic or propositional justification, but it will not matter for our purposes. I think the polarized beliefs he discusses are doxastically (in addition to propositionally) unjustified, because they are not supported by the subject's evidence; so, the question about the basing relation will not arise.

[7] One might ask whether, in saying that S *gains* or *gets* a defeater, I mean that she *discovers* a defeater that was already there independent of her awareness of it; or, whether her awareness of the defeat relationship is what *makes* the thing a defeater. On the first option, the defeater is there independently of whether it is part of S's evidential perspective. As an accessibilist, I take it this is not how Kelly conceives the matter. He would likely say that it only counts as a defeater once it becomes part of S's evidential perspective. In any case, nothing will turn on this, since I take Kelly's evidentialism on board for the purpose of this debate.

E1: [The uncongenial study presents a set of empirical results that seem to support *Not-Deterrent*], [The study's authors claim that those results support *Not-Deterrent*]
E2: [The congenial study presents a set of empirical results that seem to support *Deterrent*], [The study's authors claim that those results support *Deterrent*]
U1: [The uncongenial study contains an undercutting defeater of the evidence for *Not-Deterrent* in E1][8]

According to Kelly, then, Elena's total (new) evidence at the end of the study is: {E1, E2, U1}. And U1 defeats E1. He reasons that since Elena's total evidence is her old evidence plus these three items, with the first supporting *Deterrent*, and the rest not supporting it, her total evidence comes out more strongly in favor of *Deterrent.* So, she should become more confident. More specifically, his argument is this[9]:

(P0) Elena starts out with evidence for *Deterrent* (assumption).
(P1) The new evidence that Elena gets during the experiment is {E1, E2, and U1}.
(P2) E2 supports *Deterrent.*
(P3) E1 and U1 do not support *Not-Deterrent.*
Therefore, [from (P0), (P1), (P2), and (P3)]
(C1) Elena's total evidence supports *Deterrent.*
(P4) *Evidentialism*: You should proportion your confidence to what your total evidence supports.
Therefore, [from (C1) and (P4)]
(C2) Elena should become more confident in *Deterrent*.

4 Unpacking Kelly's normative assessment

The form of Kelly's argument is straightforward. But I will argue that it covers over important complexities having to do with what counts as part of Elena's evidence. In particular, Sect. 4 argues that Kelly needs a specific and narrow conception of what counts as part of her evidence in order to get his argument through. Section 5 argues that this narrow conception of the subject's evidence results in an implausible concept of justification.[10] On what I will argue are more plausible conceptions of what

[8] In the next section, I explain what an *undercutting defeater* is, and make a further distinction that helps narrow in on the type of defeater here.

[9] Kelly does not put his argument in premise-conclusion form; this is a reconstruction.

[10] It is worth noting that another way of arguing against Kelly would be to point out that his evidentialism does not make room for a concept of *epistemic responsibility*, and one might think that a lack of such responsibility during belief formation undermines the justificatory status of resulting beliefs. For present purposes, I bracket the question of whether epistemic irresponsibility can undermine the justificatory status of one's beliefs. Kelly has argued elsewhere that questions about how much time or effort one should devote to scrutinizing a given piece of evidence are *practical* (that is, non-epistemic) questions, so he would not be impressed by this line (Kelly 2008b, p. 13).".

counts as part of the subject's evidence, the polarized beliefs at the end of the Lord et al. study are not proportioned to the subject's total evidence, so are unjustified. As will become clear, the problem is not with how subjects responded to evidence from the uncongenial study, but rather with their taking the results of the congenial study on board at face value and increasing their confidence.

4.1 A preliminary note on why it matters

Here, I will say something preliminary about why it matters whether evidentialism counts these polarized beliefs as justified; I say more in the final section of the paper. Consider—on the concept of *justification* that counts such beliefs justified—how little it takes for a subject to justifiably become very confident that *Deterrent* is true. The subject need only justifiably believe *Deterrent*, and then be exposed to evidence that seems to support its truth. The key point is that, given how uneven scrutiny works, it does not matter whether that evidence is *good*. Since it seems to accord with the subject's antecedent view, she will take it on board without scrutiny.

According to Kelly's analysis, the resulting belief is justified. By iterating this process, one could justifiably become *very* confident that *Deterrent* is true, on the basis of a lot of bad evidence that seems at face value to support it. This result is highly counterintuitive, which suggests that our ordinary concept of justification is different from the evidentialist concept that accommodates these polarized beliefs. This should not be surprising. In ordinary contexts, judging that another person's belief is justified amounts to a kind of endorsement. And—particularly in cases where the subject is very confident—granting such endorsement seems incompatible with knowing that the belief is based on a shallow consideration of bad evidence. Other things equal, the evidentialist should want to avoid a concept of justification that gives this result.

4.2 Introducing the hard question about accessibilism

It was crucial to Kelly's setup that because Elena does not scrutinize the congenial study, she does not discover its fatal flaws, so *her* total evidence does not contain a defeater of the evidence it seems to provide for *Deterrent*. But this covers over an important question: for the evidentialist, what *exactly* comprises a subject's total evidence? That is, what counts as part of *the evidence to which the subject should proportion her belief*?

Kelly is not explicit about this, but the conception of *evidence* that his brand of evidentialism has is mind seems to be an *accessibilist* one.[11] There are many versions of accessibilism. In rough terms, it is the view that something counts as part of a subject's evidence only if it is actually or potentially accessible to her by

[11] His argument relies on the idea that the flaws in the congenial study do not count as part of Elena's evidence because she does not have access to them.

introspection or reflection.[12] But there are more and less restrictive ways of understanding what counts as *potentially accessible*. Accessibilist versions of evidentialism should therefore say something about how restrictive they mean to be, since different understandings differ substantially over *how much is required of a subject* in order to count as having proportioned her beliefs to the accessible evidence. For instance: Is S's accessible evidence at a given time simply comprised of how things evidentially *seem* to her when she turns inward at that time? Does it include things that are stored in memory, but that might take a moment to recall? Does it include all of the evidential support relations she would be able to appreciate upon a moment's reflection? Does it include all of the logical consequences of things that are currently before her mind, even if it would be difficult, or perhaps beyond her current ability, to appreciate those support relations?

Call this question of what counts as accessible evidence *the hard question* about accessibilism. Possible answers to the hard question fall on a spectrum. An answer at the most restrictive end says that a subject has potential access to a piece of evidence only if she can access it by simply turning inward and observing what is immediately given to her, so to speak. In rough terms, we can think of this as a *perceptual* model of introspection and reflection: when a subject turns inward, certain evidence is immediately available in the way that when she opens her eyes, the way things look to her is immediately available. By contrast, an answer at the least restrictive end of the spectrum might say that something counts as potentially-accessible evidence if the subject *could* access it upon very effortful deliberation, perhaps even with outside help.

The answers at both extreme ends of the spectrum are implausible, because they make justification near-automatic, or almost impossible, respectively. On the restrictive end, there is little (if any) room for the possibility of *unappreciated evidence*, since if the evidence was immediately present, the subject would (almost) certainly have appreciated it. So access*ible* evidence becomes synonymous with evidence that is in fact access*ed*. This makes justification nearly (if not entirely) automatic, since it is hard to see how a subject could fail to proportion her belief to the evidence immediately present to her. In the maximally unrestrictive case, evidence might include even that which the subject could only access in some not-so-nearby worlds. This seems to require too much of the subject in order to count as having proportioned her beliefs to the accessible evidence. Owing to the implausibility of both extremes, accessibilists should want to land somewhere in between, on a theory of justification that meshes with most of our ordinary intuitions.[13]

[12] There is a further question of whether to interpret this in a way that entails that all pieces of evidence are mental items. Accessibilists would traditionally say that a subject's evidence consists solely of mental items, but there is room in logical space for the possibility that one actually or potentially become aware of external items, such as facts about the world, via introspection or reflection. Kelly seems open to this possibility in some of his other writings. See Kelly (2008b, c).

[13] Conee and Feldman take up this issue with respect to whether or not all of the evidence stored in memory should be considered accessible. They come to a similar conclusion that the extreme views that either *none* or *all* stored evidence should count as accessible are burdened with implausible consequences, so one should want to land somewhere in between. (Conee and Feldman 2008).

4.3 Kelly on the hard question

Kelly does not say exactly what counts as potentially-accessible evidence, though his argument turns on how we define it. In the remainder of Sect. 4, I argue that Kelly's argument imposes two restrictions on what counts. The second bears more explanation and defense than the first. In Sect. 4.3.1, I introduce the first restriction and explain why Kelly needs it. In Sect. 4.3.2, I introduce the second restriction. In Sects. 4.4–4.6, I further explain the second restriction and argue that Kelly needs it in order for his argument to go through. In Sect. 5, I argue that these restrictions result in an implausible concept of *justification*.

4.3.1 First restriction

(P1) of Kelly's argument asserts that E1, E2, and U1 comprise the total new evidence that Elena gets during the Lord et al. experiment. The fatal flaws in the congenial study do not count as part of her accessible evidence. Recall that for Kelly, they do not count because why and how a subject gets her evidence is not epistemically relevant—so in Elena's case, it does not matter that she processed the information in the congenial study in a biased manner. Nonetheless, the study's flaws are in some sense potentially accessible to her, since she would uncover them if she were to process the study's information in a less biased manner. So, the *first restriction* Kelly's argument implicitly places on the definition of *potentially-accessible evidence* is that it excludes evidence the subject could have had access to, had she processed information in a less biased or unbiased manner. If this evidence counted as potentially-accessible, then (P1) would be false, as Elena's accessible evidence would also include a defeater for E2.

4.3.2 Introducing the second restriction and the MDC

The *second restriction* that Kelly's argument imposes on what counts as potentially-accessible evidence is more complex. To put it in general terms, the restriction excludes certain evidence that Elena would be able to access upon further reflection on the evidence she already has. To explain this more precisely, let me introduce a distinction that Kelly makes between two different types of evidence that subjects can have for their beliefs: *narrow* evidence and *broad* evidence. *Narrow* evidence consists of relevant information about the world—things that it would be natural to call *data*. *Broad* evidence includes evidence in the narrow sense, plus anything else one is aware of that makes a difference to what she is justified in believing. Kelly points out that broad evidence thus includes things like the space of alternative hypotheses of which one is aware.[14]

[14] Kelly makes this distinction in order to argue that if two subjects have the same narrow evidence, they might nonetheless differ in what they are justified in believing, since they have different evidence in the broad sense (Kelly 2008a).

The information Elena is given about the congenial and uncongenial studies is narrow evidence. The first restriction above thus restricts what counts as part of her narrow evidence. The second restriction restricts what counts as part of her broad evidence, holding narrow evidence fixed. That is, the second restriction excludes evidence that Elena could access by further reflecting on the narrow evidence that is already within her ken, rather than evidence she could gain by getting more information from the external world.

I will argue that without this second restriction on what counts as potentially-accessible evidence, Elena's evidence would include a defeater of at least some of the support that E2 lends to *Deterrent,* because it is within her abilities to grasp such a defeater solely by further reflection on the evidence already within her ken. To pinpoint the type of defeater I have in mind, let me introduce some terminology. First, an *undercutting* defeater is one that undermines the evidential support relation between one's evidence for a proposition, and the truth of that proposition.[15] The evidence thus loses its status *as* evidence for that proposition. In our case, for instance, U1 is an undercutting defeater of E1 because it undermines the evidential support relation between E1 and *Not-Deterrent*. Or, equivalently, U1 undermines E1's status as evidence for *Not-Deterrent.*

For Elena, U1 also acts as a *psychological defeater* of E1. Following Jennifer Lackey, I take a *psychological defeater* to be one that in fact acts as a defeater in the subject's psychology, such that the defeated evidence loses its status as evidence for the relevant proposition from the subject's subjective perspective. As Lackey puts it, "A psychological defeater is an experience, doubt, or belief that is had by S, yet indicates that S's belief that *p* is either false or unreliably formed or sustained."[16] Lackey contrasts these with *normative defeaters*, which are parallel experiences, doubts, or beliefs that S *ought* to have, given the presence of certain available evidence. For my purposes, I instead contrast psychological defeaters with what I will call *motivated defeaters*. Let a *motivated defeater* be a piece of broad evidence that (i) S could psychologically grasp solely by further reflection on her current evidence, (ii) S does not now grasp because she lacks the proper motivation; (iii) would act as a psychological defeater if grasped. Given (iii), motivated defeaters, like psychological defeaters, are experiences, doubts, or beliefs, which would indicate that S's belief that *p* is either false or unreliably formed or sustained.

Given this definition, call the specific claim I will argue for through the remainder of Sect. 4 the *Motivated Defeater Claim* (*MDC*). The *MDC* says that there exists an undercutting, motivated defeater of at least some of the evidence E2 that Elena's polarized belief is based on at the end of the Lord, Ross, and Lepper (LRL) study. Were she to become differently-motivated, and to grasp this undercutting defeater, then—at least to some degree—the evidence on which her polarized belief is based would be psychologically defeated.

Establishing the *MDC* will suffice to show that Kelly needs the second restriction on what counts as potentially-accessible evidence. If the motivated defeater were

[15] This definition is due to Pollock (1986).

[16] Lackey and Sosa (2006, p. 4).

to count as part of Elena's evidence, then, since it defeats part of the evidence her polarized belief is based on, that belief would not be proportioned to her evidence, and would be unjustified.

4.4 Unpacking the MDC

Before arguing for the *MDC*, I need to say more about the content of E2, to make clear what it would take for it to be partly or wholly defeated. Both components of E2 involve the proposition that <the congenial study's empirical results support *Deterrent*>. These empirical results are a set of straightforward empirical findings, and the idea that they support *Deterrent* implies that the fact that the results were *R* lends credence to the truth of *Deterrent*. To make clear how this works, here is an example of the actual results that subjects were shown during the LRL study:

> Kroner and Phillips compared murder rates for the year before and the year after adoption of capital punishment in 14 states. In 11 of the 14 states, murder rates were lower after adoption of the death penalty. This research supports the deterrent effect of the death penalty.

Here, the straightforward empirical results in *R* are that Kroner and Phillips compared murder rates for the years described in the states described, and that in 11 of 14, those rates were lower after adoption of the death penalty.[17] The more contentious part is the claim that *R* supports *Deterrent*.

Support is something that can come in degrees. It is also relative to a subject's evidential perspective. *Deterrent* does seem like the sort of thing that is likely to account for *R*. But other factors might also contribute, such as confounding variables, methodological issues, and the like. What determines the extent to which *R* supports *Deterrent* from a given subject's evidential perspective is the degree to which that subject has reason to think that other factors could plausibly contribute to explaining *R* (and, conversely, the degree to which they have positive reason to think that *no* such factors do). In general terms: *R* lends *some* credence to *Deterrent* so long as, given a subject's evidence, *Deterrent* is likely to account for *R*—or at least to make a substantial contribution.

E2 is made up of the fact that the study's authors *claim* that *R* supports *Deterrent*, and that it *seems* to Elena that *R* supports *Deterrent*. What do the content of this claim and this seeming amount to? The study's claim means that from its authors' evidential perspective, the truth of *Deterrent* is likely to be what accounts for *R*, or to be a substantial contributor.[18] Whether the authors have taken proper account of

[17] More specifically, let *R* stand for the fact that the researchers followed this procedure, and got this result.

[18] To put this in terms of Bayesian confirmation theory: *R* supports *Deterrent* just in case, given everything that is within Elena's evidential perspective, the prior probability of *Deterrent* conditional on *R* is greater than the prior unconditional probability of *Deterrent*. I have articulated a description of what would make this true in our case. For further discussion from the Bayesian perspective, see Earman (1991) and Fitelson (1999). But for present purposes, the intuitive description is sufficient.

everything that is in their evidential perspective is a further matter, but the act of making this claim this context at least involves purporting to have taken account of available evidence. The other part of E2—its seeming to Elena that *R* supports *Deterrent*—is somewhat more inchoate. It means that from her evidential perspective, the truth of *Deterrent* seems likely to be what accounts for *R* (or at least a substantial contributor). But being a *seeming* and not a *claim*, it does not involve a purporting on Elena's part to have taken account of everything within her evidential perspective.

Like support, defeat of E2 can also come in degrees. To the extent that Elena gains access either—(1) to evidence that the congenial study's authors may have been wrong to claim that *R* supports *Deterrent*; or, (2) to evidence indicating that *R*'s having seemed to support *Deterrent* was merely illusory—the support relation between those parts of E2 and the truth of *Deterrent* is undermined. The *MDC* says that there exists an undercutting, motivated defeater of at least part of E2. Having unpacked the *MDC*, I will now argue for it.

4.5 Arguing for the MDC

In arguing for the *MDC*, I will assume that Elena sees the congenial study before the uncongenial one, since this is the harder case to argue for: in this case, she gets a retrospective defeater of evidence she has already seen, as opposed to already having the defeater before her mind when she encounters the evidence.[19] In this case, Elena looks at the congenial study, takes it at face value, and boosts her confidence in *Deterrent*. Then she looks at the uncongenial study, scrutinizes it, and finds fatal flaws. She thus discovers that it is what I will call an *evidential* dud, meaning that it does not give her any evidence in favor of *Not-Deterrent*. She now knows that any initial seeming that its empirical results supported *Not-Deterrent* was illusory, and the authors' reasoning for that claim was fatally flawed.

Although she does not know that the congenial study is also an evidential dud, the *MDC* says that Elena has a motivated defeater of at least part of E2. In Sect. 4.5.1, I will say what the defeating evidence consists in. In Sect. 4.5.2, I will argue that it meets the characteristics that define a motivated defeater. Again, establishing the *MDC* will show that Kelly needs the second restriction on what counts as accessible evidence, for the reasons I gave earlier. Importantly, establishing the *MDC* only shows that there *is* a motivated defeater. It does not yet show on *independent* grounds that the evidentialist ought to place the first and second restrictions on what counts as evidence. I argue for that further claim in Sect. 5.

4.5.1 Defeating evidence

At the end of the LRL study, Elena has a lot of information. First, she has information about the materials she was given during the study. Subjects were given detailed

[19] In the LRL study, half of the subjects see the congenial study before the uncongenial one, and half see the uncongenial one before the congenial one.

descriptions of the purported research they were asked to evaluate, including information about procedures and methods, explanations of criticisms of the studies in the literature, and authors' rebuttals to those criticisms. They were instructed to use their evaluative powers to think about what the studies did, what the critics had to say, and whether the responses to those criticisms were adequate. So, Elena has certainly been given reason to believe that there is a live scholarly debate among experts about whether the congenial study provides adequate evidence for its purported conclusions. It would thus be evidentially irrational to take it on faith that the responses adequately diffuse criticisms, without making an effort to understand the purported merits of the criticisms, and how the replies defeat them.[20]

Having read all of this, subjects were asked to rate the study's convincingness, and to write a description of why it did or did not support *Deterrent*. This writing task forces them to reflect on—or at least take stock of—the evidence they have gained during the course of the study. This provides Elena another opportunity to access the asymmetry between her reasons for dismissing the uncongenial study, and for accepting the congenial one, by putting those reasons side by side. Indeed, subjects' actual written responses from the LRL study make the asymmetry quite explicit. For instance, subject S8 writes of the congenial study, "It does support capital punishment in that it presents facts showing that there is a deterrent effect and seems to have gathered data properly." This evinces acceptance of the congenial study's results, but does not show that S8 understands purported criticisms of the study, or has reasons for thinking them invalid. Of the uncongenial study, S8 writes, "The evidence given is relatively meaningless without data about how the overall crime rate went up in those years." Unlike the previous statement, this evinces understanding of the criticisms of the uncongenial study. Although this asymmetry may not act as a psychological defeater because it is not evident from the subject's first-person perspective, I argue in Sect. 4.5.2 that it is part of a motivated defeater or (at least some of) the subjects' evidence in E2.

In addition to all of this first-order and meta-information about her understanding of the congenial study, Elena also has information about the relationship between the different pieces of her new evidence. Recall that both the congenial and the uncongenial study's results are of the form "*R*. *R* supports *p*."[21] where *R* is the kind of thing that could explain *p*, and that *prima facie* seems to support *p*. When Elena grasps U1, she then has, occurrently before her mind: (1) an *example* and an *understanding* of a way in which a *prima facie* plausible claim of the form "*R* supports *p*" can turn out to be defeated, and (2) an *experiential understanding* of how the defeaters of such claims may be non-obvious, since she had to scrutinize in order to find U1. I will argue that this also contributes to the motivated defeat of E2. Were Elena differently motivated, this understanding of how a structurally-identical claim can be

[20] Presumably, Kelly would reply to this charge of evidential irrationality by pointing out that it might still *seem* to Elena that she understands all of this (though in fact she does not, since some of the criticisms name fatal flaws in the studies). For Kelly, that seeming is part of the broad evidence to which her belief is proportioned. Section 4.5.2 will explain why this does not undermine the argument for the *MDC*.

[21] Here, of course, *p* represents *Deterrent* and *Not-Deterrent*, respectively.

non-obviously defeated would raise the possibility of defeat to salience with respect to E2 as well. Thus, Kelly's inference from [(P0), (P1), (P2), and (P3)] to (C1) does not go through because we cannot "read off" a fact about what Elena's total evidence supports from facts about what individual pieces of it support. The individual pieces of evidence bear on each other in ways that are not captured by looking summatively at the individual support relations.

4.5.2 Why this comprises a motivated defeater: an empirical argument

This section argues that the information outlined in Sect. 4.5.1 comprises an undercutting, motivated defeater of at least some of the evidence in E2. Start by considering how one might argue that it is *not* within Elena's abilities to understand the information within her ken as a defeater of E2. One might think, for instance, that Elena's beliefs are compartmentalized in a way that would prevent her, even upon active introspection and reflection, from accessing this defeat relation.[22] While I grant the *possibility* that an epistemic agent could be such that her beliefs are compartmentalized in this way, I will present evidence from the psychological literature that we are not generally like this.[23] That is, I draw on empirical literature to show that it is not beyond Elena's *ability* to understand the information within her ken as a defeater of E2; rather, the problem is that she is motivated to resist understanding the information this way. In this sense, she *could* access a defeater of E2 simply by further reflecting on the evidence within her ken. Kelly needs to exclude such evidence from counting as potentially-accessible, via the second restriction, in order for Elena's polarized belief to count as justified at the end of the LRL study. In Sect. 5, I will make the further normative argument that the evidentialist has independent reason to forego this restriction.

The *heuristic-systematic model* of information processing *(HSM*) is widely recognized in psychology as a model of how people process information that is relevant to their beliefs during inquiry. The model distinguishes between different kinds of motivations that people can have while processing information (Chaiken et al. 1996). It distinguishes states in which the subject's primary motivation is to arrive at an accurate, well-founded, and unbiased understanding of the matter about which they are inquiring (*accuracy motivation*), from states of *goal-oriented* information processing, in which the person's primary motivation is either to defend their pre-existing beliefs, worldviews, or self-concepts (*defense motivation),* or to make a

[22] Thanks to Zoe Jenkin for pointing out the possibility of this compartmentalized subject to me.

[23] Would this hypothetical agent's polarized belief count as evidentialist-justified? This depends on how we answer the hard question, since it depends on just how much unsuccessful introspection or reflection is allowed before the agent counts as "not having access" to the defeat relation. Some accessibilists (see Ginet 1975, p. 34; quoted in Alston 1989, p. 213) *do* insist that the evidence has to be "directly recognizable" in order to count as accessible, but I do not find this plausible since it makes justification nearly automatic. Supposing the defeat relation is not accessible on *any* amount of introspection and reflection, I concede that the hypothetical agent's belief is evidentialist-justified. If you judge that this is the wrong result, then perhaps it is so much the worse for evidentialism. Still, I do not think the subjects that Kelly has in mind are generally like this.

desirable impression on others (*impression motivation*).[24] It also distinguishes two different modes of information processing: *systematic processing*, in which subjects thoroughly scrutinize the quality of relevant information and arguments, and *heuristic processing*, in which they rely on cognitive shortcuts. A subject's motivation, as well as their desired level of confidence, influences which type of processing they use in a given situation. In general, people will spend only as much cognitive effort as is required to satisfy their goal (*least effort principle*), and they will spend whatever effort is required to attain a sufficient level of confidence to accomplish that goal, so long as they have the capacity to do so (*sufficiency principle*).

I will focus on the distinction between accuracy motivation and defense motivation. It has been established across a number of different experimental contexts that accuracy and defense motivations influence the reasoning strategies that a subject applies in a given context of inquiry. These motivations can affect encoding, organization, and use of new information that bears on a subject's inquiry.[25]

When subjects are accuracy motivated, they choose reasoning strategies appropriate to their goal of gaining an accurate understanding of the subject of their inquiry. For instance, accuracy motivation has been found to reduce or eliminate subjects' susceptibility to cognitive biases like the *fundamental attribution error*, and *anchoring effects* in probability judgments (Freund et al. 1985; Tetlock 1985; Pittman and D'Agnostino 1985), and it has been found to lead to reduced confirmation bias in information selection (Lundgren and Prislin 1998). More generally, accuracy motivated subjects tend to process information thoroughly and cautiously, and are more likely to produce accurate judgments as a result (Freund et al. 1985). Because they process information thoroughly and systematically, they are more likely to accurately distinguish strong from weak messages (Clark et al. 2008, 2012; Hart et al. 2009).

When a subject is defense motivated, they choose strategies that allow them to defend their pre-existing attitudes. The hallmark of defense motivation is therefore a self-serving, directional bias in processing (Chaiken et al. 1996). The most obvious effect of this is that when defense motivated, subjects produce inaccurate assessments of information that conflicts with the attitudes they are motivated to defend. For instance, Liu (2017) found that defense motivated subjects rated weak arguments that were compatible with their pre-existing attitudes as stronger than they actually were, while they rated strong arguments that were incompatible with their attitudes as weaker than they were.

Interesting interactions have been found between defense versus accuracy motivation, and heuristic versus systematic information processing. It is not simply that defense motivated subjects use heuristics, while accuracy motivated subjects process information systematically. Rather, defense motivated subjects selectively process information in the way that best meets their defensive needs. In line with the

[24] Motivation here is defined as any wish, desire, or preference on the part of the subject that concerns the outcome of a given reasoning task.

[25] For a review, see Srull and Wyer (1986).

sufficiency principle,[26] those who are highly motivated to defend their pre-existing attitudes often expend greater effort, engaging in systematic processing, because they judge that it will help justify the attitudes they seek to defend. For instance, Ginossar and Trope (1987) found that defense motivated subjects used base rate information when doing so helped justify their belief. More generally, when defense motivation is high and cognitive resources are available, defense-motivated systematic processing is likely to emerge, characterized by effortful but biased scrutiny and evaluation of judgment-relevant information (Chen and Chaiken 1999). As in the LRL study, subjects in these conditions are likely to judge congruent information more favorably than incongruent information (Pomerantz et al. 1995; Pyszczynski and Greenberg 1987), and to engage in systematic processing in order to subject the incongruent information to greater scrutiny, and undermine its validity (Ditto and Lopez 1992; Giner-Sorolla and Chaiken 1997; Liberman and Chaiken 1992). Numerous studies have thus found that people are more sensitive processors of information they do not want to believe than of information they do want to believe (Ditto et al. 1998).

Correspondingly, congruent with the least effort principle, defense motivated subjects are not likely to engage in systematic processing of information that appears congenial to the attitudes they wish to defend. Rather, they are likely to take it at face value. Gawronski and Bodenhausen (2006) explain that if a subject's automatic affective response to new information aligns with their preferred attitude, then the search for additional relevant information may be truncated, and evaluative judgments based largely on affirmation of the automatic affect (Baumeister and Newman 1994; Ditto and Lopez 1992).

Given all of this, we are in a position to see that subjects in the LRL study—as well as in the Houston and Fazio (1989) study that replicated it—exhibit a reasoning profile typical of defense motivation. In line with the sufficiency principle, they engage in systematic processing of information that appears uncongenial to the view they wish to defend, subjecting it to greater scrutiny in order to undermine its validity. And in line with the least effort principle, their evaluation of information that appears congenial to their view is based largely on affirmation of their initial, positive affective response.

A further reason to believe that subjects in the LRL study are defense motivated is that the setup of the LRL study involves factors that have been found to increase defense motivation. Personal commitment to an attitude or belief increases defense motivation, where such commitment may be caused by one's having freely chosen the view, without coercion (Hart et al. 2009), as subjects did at the outset of the LRL study. Harmon-Jones and Harmon-Jones (2008) explain that when a view is freely-chosen, there is a need to make alternatives seem less attractive in order to reduce the unpleasant emotion of cognitive dissonance. This describes what subjects in the LRL study did. Secondly, defense motivation has been found to arise with respect to self-definitional attitudes and beliefs (Chaiken et al. 1996), which are attitudes and

[26] Note that when subjects are defense motivated, sufficiency is determined by whether processing yields a judgment that reinforces the attitudes one is seeking to defend with the desired level of confidence.

beliefs that for example involve one's values and social identities (Chen and Chaiken 1999). One's views about the usefulness of the death penalty plausibly involve both. In these situations, defense-motivated subjects process information selectively in order to preserve their self-concept and associated world views.

Defense and accuracy motivations can be manipulated experimentally. Numerous methods have been found to increase accuracy motivation in experimental subjects. For instance, accuracy motivation increases if subjects anticipate having to explain the basis of their judgments to others (Chaiken 1980; Freund et al. 1985; Kunda 1990, 1999; Leippe and Elkin 1987; Petty and Wegener 1999; Tetlock and Kim 1987), or if they are told that their reasoning abilities will be evaluated (Lundgren and Prislin 1998). Outcome-relevant involvement has also been found to foster accuracy goals. For example, Jonas and Frey (2003) induced accuracy motivation by telling participants that they would receive a prize for a correct choice.

The upshot is that subjects like Elena have the *ability* to access the partial defeat relation between E2 and the evidence already within their ken. They simply lack the *proclivity* to do so, given their current defense motivation. And it is well established that such motivations can be manipulated with experimental interventions. In fact, Schuette and Fazio (1995) performed an experiment in which they used Lord et al.'s original paradigm, but induced accuracy motivation in half of the participants at the beginning of the experiment by telling them that their judgments about the studies would be compared to the judgments of an expert panel of "eminent social scientists that recently had evaluated research on capital punishment, including the two target studies (Schuette and Fazio, p. 707)." Participants were also told: (1) that both studies had received clear and unanimous judgment from the panel, and that the purpose of the experiment was to judge whether laypersons could match the correct answer provided by experts; and, (2) that there would be a brief discussion afterwards of why, in the subject's view, their judgments did or did not match those of the panel. The result was that subjects in this condition showed no relation between their attitudes and their judgments, while subjects in whom accuracy motivation was not induced, like those in the original LRL study, readily accepted the interpretation implied by their antecedent belief. The researchers concluded that "High fear of invalidity apparently motivated subjects to expend the effort to consider the value of the attributes more thoroughly and objectively, instead of simply accepting the interpretation implied by their attitudes (p. 710)."

The results of this experiment do not bear directly on the *MDC*, since instead of looking at the evidence subjects actually have at the end of the LRL study, it shows that their evidence would be different if they had been accuracy motivated from the beginning, because they would not have engaged in uneven scrutiny. To provide evidence for the *MDC*, subjects would have to be told only after having finished looking at all of the information about the congenial and uncongenial studies that their evaluations of these studies would be judged by experts. What this study *does* provide is an example of how subjects have the ability to access evidence that they may not be motivated to access—or may be positively motivated *not* to access.

My positive suggestion will be that a more moderate and reasonable answer to the hard question should be constrained by facts about what subjects have the ability to access without too much scaffolding, rather than by their proclivities given their

current motivation. Otherwise, accessibilist evidentialism has the unintended effect of making it the case that what one is justified in believing turns on one's current motivations. In Sect. 5, I will go into more detail about why the restrictive conception of evidence needed to make Kelly's argument work is problematic. Specifically, I argue that the evidentialist has independent reason for answering the hard question in a way that places neither the first nor the second restriction on what counts as potentially accessible evidence. Otherwise, the evidentialist's picture of justification is untenable, and out of sync with our ordinary concept. Before making the normative argument of Sect. 5, let me address a few objections to the idea that it would be evidentially rational for Elena to understand the evidence I outlined in Sect. 4.5.1 as a partial defeater of E2.

4.6 Objections and replies

4.6.1 Bayesian belief polarization

What it is rational for Elena to believe at the end of the LRL study depends on the evidence that she came in with, and on whether she combines it with her new evidence in a way that is consistent with normative principles of belief revision. A number of authors have shown—using Bayesian networks to characterize different kinds of relationships between hypotheses and data—that some instances of belief polarization are consistent with a normative account of belief revision.

An anonymous reviewer helpfully points out that a person who uses Bayes' rule counts as an evidentialist on Kelly's notion of evidentialism. Thus, there is an evidentialist-rational route to polarization. And Kelly's conclusion—that subjects' polarized beliefs at the end of the LRL study are justified—is warranted for subjects who use Bayes' rule to arrive at their polarized beliefs. The question this leaves us with is whether the kind of uneven scrutiny that subjects like Elena engage in is part of a process of using Bayes' rule to revise their belief in response to new evidence. In what follows, I briefly summarize some of these models as they apply to the polarization that happens in the LRL study, and then I argue that this is likely not how most subjects in the LRL study are actually reasoning.

A number of authors have argued that polarization can arise as a result of rational belief revision processes.[27] I will focus on a set of arguments from Jern et al. (2014), since they address the claim that the polarization that results from the LRL study in particular is consistent with a normative account of belief revision. The important general point they make is that belief polarization can result from normative probabilistic inference in accordance with Bayes' rule when subjects make different assumptions about factors that affect the relationship between hypothesis H and data D. They use Bayesian networks to model situations where a third variable V

[27] See Benoît and Dubra (2019), Gerber and Green (1999), Glaeser and Sunstein (2013) and Jern et al. (2009, 2014).

affects the outcome of D.[28] They then use this to develop normative accounts of the polarization that occurred in the LRL study, giving two specific explanations of how subjects' beliefs might have resulted from normative inference. Given space restrictions, I will only go through the first of these, as the main part of my response below would be the same for both.

The first explanation is based on two assumptions that a participant might have made: (1) studies like the ones that participants read about are influenced by bias such that researchers tend to arrive at conclusions that are consistent with their own prior beliefs; and, (2) that one's own beliefs about the effectiveness of the death penalty differ from the consensus opinion among researchers and other experts. Jern et al. compare two hypothetical participants: Alice, who initially believes *Deterrent*, but thinks that her belief is the opposite of the consensus expert opinion, and Bob, whose beliefs are the opposite of Alice's. They compute Alice and Bob's updated beliefs, conditioning on the data that one of the studies Alice and Bob were shown supports *Deterrent*, and the other supports *Not Deterrent*. They conclude that given her assumptions, Alice's prior belief in *Deterrent* provides her with a justification for treating the study supporting *Not Deterrent* as a spurious result due to researcher bias, so she becomes more certain of her antecedent belief (as does Bob, for the same reason). It is worth noting here that this cannot be the way that subjects in the actual study reasoned, because Lord et al. only used subjects—from a survey administered earlier—who thought that most of the relevant research favored their position.

Considering Alice and Bob merely as hypothetical participants, then, I will give two responses. The first concerns the rationality of these imagined subjects' belief updating processes, and the second is a more fundamental point that applies to all of the arguments in this literature that aims to show that belief polarization can result from normative probabilistic inference in accordance with Bayes' rule. First, it is not clear whether Alice should take the fact that one of the studies she is shown seems to offer support for *Deterrent* to evince that the authors of that study were unbiased because they got that result despite not antecedently believing *Deterrent* themselves. Instead, perhaps she should think that the authors of this study are part of the minority of researchers and experts who antecedently agreed with her, and that their result would also be influenced by bias, given their motivation to show that the majority view is incorrect.[29] This depends among other things on how small she thinks the minority of researchers and experts that agree with her is. If she thinks it is very small, then she may think it unlikely that the authors of the congenial study are part

[28] More specifically, they model cases in which: V is an additional factor that bears on H; V informs the prior probability of H; D is generated by an intervening variable V; V is an additional generating factor of D; H and D are both effects of V; and, V informs both the prior probability of H and the prior probability of D.

[29] One might object here that since Alice believes *Deterrent* is true, she should believe that studies whose results support *Deterrent* get those results *because Deterrent* is true. But this assumes that all studies that get a true result are well-designed, and that the researchers' motivation to show that their minority view is actually the correct one does not bias their research designs.

of that minority. But then we might worry about the rationality of her disagreement with what she takes to be a near-consensus view among researchers and experts.

A second, and more fundamental response applies not just to this specific explanation from Jern et al. but in general to the Bayesian arguments showing that belief polarization can be consistent with normative probabilistic inference. Fundamentally, their point is not inconsistent, or even in tension with mine. They show that there are rational routes to belief polarization. My argument does not bear on whether this is true. I am only concerned with the particular belief revision processes that produced the polarized beliefs that subjects like Elena had at the end of the LRL study.

Jern et al. are quick to point out that: "Our two Bayes net accounts of the death penalty do not imply that participants diverged in this study for normative reasons… We do not claim that either account is the correct explanation for the study results. We propose only that these accounts are plausible explanations that cannot be ruled out a priori (pp. 212–213)." They are of course right that such accounts should not be ruled out a priori. Importantly, my argument does not reason from the fact that polarization occurred to the conclusion that biased reasoning took place. The polarization itself does not yet tell us how subjects are reasoning. Rather, it is the empirical argument for the *MDC* that I gave in Sect. 4.5.2 that demonstrates the improbability that subjects in the LRL study were using Bayes' rule to revise their beliefs. To see why, first note that Bayesian networks are designed to model how subjects *should* reason in order to maximize their changes of arriving at true beliefs under conditions of uncertainty. We should therefore expect subjects to reason this way—assuming they are able to—only insofar as they are both accuracy motivated and instrumentally rational. By contrast, it would not be instrumentally rational for a defense motivated subject to reason this way, since it risks proving her antecedent belief wrong. As explained in Sect. 4.5.2, defense motivation is a determining factor for how people reason under conditions of uncertainty. And as I argued there, we have good reason to believe that subjects in the LRL study were defense motivated. Given all of this, we have two kinds of reasons to believe that subjects in the LRL study did not reason according to Bayes' rule. First, though epistemically rational, it would not have been instrumentally rational for them to do so given their defense motivation. Second, there are the empirical reasons: in addition to those that I detailed in Sect. 4.5.2, there is the fact that the LRL subjects' written testimonies do not evince this kind of reasoning.

4.6.2 The background belief objection

A related objection starts by asking us to suppose Elena came in with a justified belief that it is very rare for these types of empirical studies to be evidential duds.[30] Upon scrutinizing the uncongenial study, she finds that it is a dud. The fact that at least one of the two studies she has been shown is a dud then becomes part of her total evidence. Evidential rationality requires her to conditionalize on this new knowledge, which should lower her confidence that evidential duds are quite rare, given that one of the two studies she has been shown is a dud. She ought to believe there are more duds out there than she previously thought, and to boost her confidence that any given study, including the congenial one, is a dud.

So far this is not controversial. However, the objector points out that there is an open question about *how much* Elena ought to lower her credence that duds are very rare and boost her credence that the congenial study is a dud. And on a reasonable way of answering that question, the objector says, she ought not change her credence very much, because the discovery of one more dud is very weak statistical evidence that E2 is also a dud. I will offer two replies. The second is based on a further specification of the objection.

The objector is right that how much Elena ought to change her credence depends on a variety of different factors. For one, it depends on how confident she was in her initial belief that duds are very rare. It also depends on how she ought to understand the relevant class of *these types* of studies. Presumably, the objector is thinking that she should conceive of these types of studies quite broadly, perhaps as any that aim at settling a complex empirical matter using relevant data. My first reply is that this is arguably not how she ought to conceive of the relevant class of studies. Given her initial belief that duds are rare, the discovery of a dud should be surprising. Arguably, this surprising discovery is best explained by the idea that the studies she is being shown are not a random sampling. So, she should understand *these types* of studies as *studies that I am being shown in the context of this experiment*. The two studies subjects are shown also make structurally identical claims, which provides further evidence of curating.

The objector might not accept this first reply. Indeed, there is room for reasonable debate about how Elena ought to conceive of the relevant class of studies, and more generally, about how much she ought to lower her credence that duds are rare upon discovering U1. The objector might press that on reasonable ways of answering these questions, it turns out that Elena is still rationally permitted to believe that duds are somewhat rare. And, so long as she is rationally permitted to believe that *more than 50%* of these studies are non-duds, she may end up with a justified polarized belief. Let me explain. If she justifiably believes that more than 50% are non-duds,

[30] In Bayesian terms: her prior probability distribution for how often studies of this sort are evidential duds is skewed heavily towards its being very rare. Note that this is a charitable supposition, since if she started off believing duds are somewhat common, then it was irrational in the first place for her to take E2 at face value without first scrutinizing the congenial study to rule out the possibility of its being a dud. Note also that this does not address subjects who *know* the percentage of duds in advance. Such subjects are presumably very rare.

then *even if* she has no special reason to believe that the congenial study in particular is a non-dud, she may still justifiably believe it is more likely than chance to be one. And conditional on there being a study that has a greater-than-50% chance of containing genuine evidence of *Deterrent*, her credence in *Deterrent* should go up, even if only a little. So, the objector says, when Elena leaves the study, her credence in Deterrent should still be higher than it was when she came in. This would mean her belief is both polarized, and epistemically rational.

This seems right, so far as it goes. It may thus turn out that at the end of the study, Elena's credence in *Deterrent* ought to be slightly higher than it was when she came in. This will depend on the factors I mentioned earlier, as well as on how reliable she thinks non-duds are (studies lacking fatal flaws may still not be 100% reliable). But, importantly, this route to polarization is different from the one that Kelly describes. When Elena boosts her credence in *Deterrent*, she is not doing the statistical reasoning just described. Rather, she is feeling the force of not having found anything wrong with the congenial study and treating that as evidence that the congenial study in particular is a non-dud. This is evinced by the fact that when asked, subjects expressed the belief that *the congenial study was convincing*.[31] They did not express the belief that given their total evidence, the congenial study had a greater-than-50% probability of being a non-dud, and therefore provided *some* evidence in favor of their antecedent belief, even without scrutinizing it to see whether this probable state of affairs matches the actual one. So, I agree with this objector, as with the previous one, that there may *be* an evidentialist-rational route to polarization, but it is not the one Kelly describes.

4.6.3 The scope objection

In my first reply to the previous objection, I argued that Elena might have evidential reason to consider the merits of the studies she is being shown in this context apart from the larger pool of empirical studies out in the world. One might point out that this reply results in the argument of this paper having limited application to the more general phenomenon of belief polarization that is *not* generated by evidence delivered in the laboratory, in a single setting, or in structurally-identical form.

This is not an objection to my argument, but an observation about its scope. I will make two points in response. First, in terms of drawing definitive conclusions about particular cases, the scope of my argument *is* limited: it applies only to the particular routes to polarization that Kelly targets. But my argument for the *MDC* also gives us the resources to explain *why* we have good reason to proceed with caution in drawing definitive conclusions about belief polarization in general. Namely, drawing such conclusions depends on being able to give a general answer to the hard

[31] Of the congenial study, subjects said things like, "It shows a good direct comparison between contrasting death penalty effectiveness. Using neighboring states helps to make the experiment more accurate by using similar locations" and "It does support capital punishment in that it presents facts showing that there is a deterrent effect and seems to have gathered data properly". Quotes like this express subjects' beliefs that they *do* have special reason to think that the congenial study is a non-dud.

question about accessibilism. I argued that Kelly's account covers over this question, while presupposing an answer that does not seem plausible. It is a merit of my argument that it sheds light on how the hard question makes it difficult to draw definitive conclusions about belief polarization in general.

A second reply is that although I refrain from drawing definitive general conclusions about belief polarization, what I have said points the way towards certain conclusions that are more general, at the cost of being less definitive. There are two general questions we might consider about polarized beliefs that result from uneven scrutiny, beyond the context of the LRL study. The first question is whether it is evidentially rational in general to take congenial evidence on board without scrutiny. Many instances of this are evidentially irrational, insofar as they are result of a prevalent bias that psychologists call *belief bias*. Belief bias occurs when a person judges the strength of an argument based on the plausibility of the conclusion rather than on how strongly it supports that conclusion. Such judgments rely on a heuristic wherein the plausibility of the conclusion acts as a stand in for judging the strength of the argument. This is evidentially irrational, since in many cases the strength of the conclusion itself gives very little (if any) reason to think that the particular argument under consideration provides additional evidence in its favor.[32]

The second general question is whether, after repeated instances of uneven scrutiny, a subject ends up in a situation where evidential rationality requires that she be skeptical of the way things evidentially seem to her. After many instances of uneven scrutiny, a subject will end up in a situation where it evidentially seems to her that there are a lot of very convincing arguments or evidence—and no bad arguments—in favor of the proposition she believes. At the same time, it will seem to her that there are a lot of bad arguments in favor of its negation. Should this make her suspicious of how things evidentially seem to her? Again, the answer depends on how much evidential rationality requires of us in general. But on a moderate view where it requires taking account of more than just how things immediately appear a given time, it is not beyond the pale to think that a subject in this situation rationally ought to be suspicious.

4.6.4 The special evidence objection

A final objection to my argument holds that Elena need not be suspicious of the congenial study because she has evidence that the congenial study in particular is not a dud. I will consider two pieces of evidence that the objector might suggest.

The first suggestion is that Elena knows that she read the congenial study, and did not find fatal flaws. But this should not count for Elena as evidence that the congenial study is not a dud. Recall: it was not until she *scrutinized* the uncongenial

[32] In this case, the strength of the conclusion *Deterrent would* give a subject reason to believe that researchers would find patterns in the data the evince *Deterrent, if* researchers used sound methodology (sound experimental design and data collection; eliminating confounding variables; interpreting results correctly, etc.). But the strength of the conclusion does not *in itself* give subjects any reason to believe that researchers used sound methodology.

one that she was able to find its flaws. She thus knows that flaws that confer dud-status can require effortful scrutiny to uncover.[33] So, her impression that the congenial study seemed convincing is not good evidence that it did not contain such a flaw.

Second, an objector might suggest that because E2 favors a view that Elena independently believes is true, she has some extra reason to think that E2 is not a dud. This objector makes the mistake that underlies the *belief bias* phenomenon that I described previously. They fail to separate *reason to believe the conclusion* from *reason to think that the argument in favor of that conclusion is a good one*.

5 Revisiting the hard question

I argued that Kelly needs a particularly restrictive conception of what counts as part of a subject's accessible evidence in order for his argument to work. This section explains in more detail why this is problematic, by giving a more general argument against the relevant kind of accessibilist evidentialism.

There is some reason to think that at first blush, Kelly might be happy to accept a highly restrictive view of what counts as accessible evidence. On his (2008b) view, there is a fundamental distinction between epistemic and instrumental rationality, such that questions about how much effort one should devote to scrutinizing a given piece of evidence are practical questions. So it would not be surprising if he treated the question of how much *cognitive* effort to put into accessing one's evidence as a practical one, too, such that one cannot have epistemic reason to put effort into accessing evidence by introspection or reflection. For Kelly, then, the evidentialist mandate to proportion one's belief to her total accessible evidence may indeed include only evidence that is immediately accessible when the subject turns inward.

Why does it matter whether the evidentialist answers the hard question in this restrictive way that allows us to count the polarized beliefs that Kelly discusses as justified? In Sect. 5.1, I will say more about the concept of justification that this restrictive picture of accessible evidence leads to, and further explain the claim I made in Sect. 2.1 that it is out-of-sync with our ordinary concept. In Sect. 5.2, I argue that in addition, this concept cannot succeed as what I call *revisionary* and *theoretical* analyses of justification, either. In Sect. 5.3, I explain what is missing from the concept, and make a suggestion about how evidentialism might handle this.

5.1 Not our ordinary concept

We saw that in order to get his argument through, Kelly needs the first and second restrictions on the evidence to which the subject must proportion her belief. This results in a picture of justification on which a subject's belief cannot be criticized for either (i) her failing to proportion it to evidence she could have access to if she

[33] This again presupposes a moderate answer to the hard question, which I will argue for at greater length in Sect. 5. If the subject does not put two and two together here, I argue that it is not because she lacks the ability, but because she lacks the proclivity given her current motivations.

had processed information is a less biased manner; or, (ii) her failing to proportion it to certain evidence that she could access upon further reflection on the evidence already within her ken. Call this picture of evidentialist justification the *conservative* picture of evidentialist justification.

This conservative picture is out-of-sync with our ordinary concept of justification, since on our ordinary concept, (i) and (ii) might well undermine a belief's justification. Indeed, they articulate some of the things that drive our intuitive judgment that Elena's belief is unjustified. To see this, imagine a slightly different version of the case, starring *Elena2*. *Elena2* gets exactly the same new evidence that Kelly says Elena has: {E1, E2, U1}. But she acquires the evidence via testimony. Her version of the study is set up so that someone *else* reads all of the information about the congenial and uncongenial studies, and then conveys it to her. And *Elena2* is given ample reason to believe that her informant is not only maximally competent with respect to interpreting these types of studies, but also completely unbiased, and epistemically virtuous in every respect. Unbeknownst to *Elena2*, however, her informant turns out to be biased, and to fall short of epistemic virtue in this case. Indeed, the informant processes the information in the exact same manner as Elena did in the original case, subjecting it to uneven scrutiny, and failing to reflect on how different parts of her evidence bear on one another.

Intuitively, when *Elena2* ends up with the same polarized belief that Elena did, her belief (unlike Elena's) *is* justified. The crucial difference is that *Elena2*'s setup removes the possibility of criticizing her belief on the basis of (i) and (ii). Regarding (i): since *Elena2* bears no responsibility for the genesis of her evidence, there is no evidence she could have had access to if she had processed information in a less biased manner. As for (ii): *Elena2* has a defeater defeater for the defeat relationship between E2 and the evidence I described in Sect. 4.5.1, since she has every reason to think that her informant is maximally competent and unbiased. She thus has reason to think that if the congenial study contained a fatal flaw, the informant would have found and reported it to her. When we remove the possibility of criticizing the subject's belief on the basis of (i) and (ii), but keep her evidence the same, we no longer judge her polarized belief unjustified. This suggests that on our ordinary concept of justification, (i) and (ii) articulate some of the very things that can undermine justification. Thus, the conservative picture of evidentialist justification is substantially different from our ordinary notion.

5.2 Conservative evidentialism as a revised or a theoretical account of justification

Conservative evidentialism is out-of-sync with our ordinary concept of justification. But sometimes, philosophical analyses depart from our ordinary concepts, either to *revise* those concepts, or in the service of developing independent, theoretical concepts that do not relate to our ordinary ones in any straightforward way. Call the former a *revised* concept and the latter a *theoretical* one. In this section I argue that the conservative picture cannot offer either concept of justification. I first argue that it cannot succeed as a *revised* concept because it gives up on a particularly useful,

and epistemically beneficial function of the original concept. Then, I argue that it is not suitable as *either* a revised *or* a theoretical concept because it in effect *changes the subject*, and that is not something we should want our revised and theoretical accounts of normative epistemic concepts to do. Both arguments center around the claim that the conservative picture removes—to an unacceptable degree—the possibility of holding one another accountable for important aspects of the belief forming process in which our doxastic agency is involved. And this is one of the main social and cognitive functions of our ordinary concept of justification.

To see what this function consists in and how it is epistemically beneficial, consider the following. We live in an epistemic community. Since we each have limited individual faculties for collecting evidence and forming beliefs, we have an epistemic division of labor. So, to gain true beliefs about the things we care about, we often rely on the results of other peoples' belief forming processes. When a friend tells me that p is true, I believe it not because I have investigated the evidence for p independently, but because I trust that she has gathered and weighed evidence appropriately.

When we acquire beliefs in this way, other people's belief-forming processes act as a kind of stand in for our own.[34] This is an epistemically efficient division of labor, but only insofar as the testifiers are using belief-forming processes that we would also accept. Thus, we need a way of coordinating belief-forming processes, both (a) so that other peoples' belief-forming processes are generally acceptable to us as stand-ins, and (b) so we can indicate when they are not. Our ordinary concept of justification, and corresponding terminology,[35] are invaluable in this regard: when we recognize that another person's belief was formed by a process we would not accept, we can express our disapproval by labeling their belief *unjustified*. This offers epistemic criticism, indicating that the way the belief was formed or revised is unacceptable on epistemic grounds. And these criticisms influence behavior. Where we have control over the relevant aspects of our belief forming processes, such expressions have an overall tendency to influence the audience to follow the implicitly endorsed belief-forming rules and practices (and to refuse to follow those that are not endorsed). When used throughout the epistemic community, these practices serve to coordinate belief-forming processes in accordance with (a). That is, iterated use of our concept of justification in the service of (b) is a means to coordinating belief-forming processes in the service of (a).

In revising our ordinary concept of justification, we ought not give up on this function. To maintain this division of epistemic labor, we need a concept that allows us to coordinate belief-forming processes in this way. And the conservative evidentialist picture of justification gives up on the features that allow it to function this way.

[34] I take cues here from Sinan Dogramaci's work. For a detailed explanation, see Dogramaci (2011) and Dogramaci (2015). He says that our practice with terms like *rational* functions to extend our collective epistemic reach by enabling each person to serve as an "epistemic surrogate" for any other.

[35] When I refer to "corresponding terminology" here, I include not only the terms *justified* and *unjustified*, but also terms like *rational*, *irrational*, *reasonable*, *unreasonable*, and perhaps others. These are often used interchangeably to evaluate beliefs in ordinary contexts.

Next, I offer a reason to think the conservative picture of evidentialist justification does not work as *either* a revised *or* a theoretical account, because in giving up on this key function of our ordinary concept, the conservative picture *changes the subject*, which is not something we should want our revised or theoretical accounts of normative epistemic concepts to do.

When we are investigating concepts in a revisionary or theoretical mode, there are limits on *how* revisionary we can be before we end up changing the subject. Put simply: if a revised or theoretical concept departs too far from the original, then it is no longer the same concept. More specifically: if a revised or theoretical concept is unable to serve the main cognitive and social purposes of the original (or at least to serve purposes that are continuous with them), then there is not enough to ground it as being the same concept.[36] We can thus argue that the fact that conservative evidentialism counts Elena's belief as proportioned to her evidence does not show that such polarized beliefs are justified; it simply changes the subject.

I suggested that the conservative notion is unable to serve an important social and cognitive function of our ordinary concept, in allowing us to coordinate belief-forming processes, and achieve an efficient division of epistemic labor. Why think in addition that this is one of the *main* functions of our concept? Because our ordinary concept was likely shaped in response to this very need. That is, our need of a concept that could play this role in our social epistemic practice likely explains why our ordinary concept is what it is in the first place.[37]

5.3 The purpose of our normative epistemic concepts

In the previous section, I said that one of the main cognitive and social functions of our ordinary concept of justification is that it lets us to encourage one another to do better epistemically, with respect to those aspects of the belief-forming process over which we have control. This of course presupposes that we are the kinds of doxastic agents who sometimes exhibit a significant measure of control over our belief-forming processes. In this section, I argue that the conservative evidentialist notion of justification is not well-suited to evaluate the beliefs of doxastic agents like this.

[36] I borrow the general form of this thought about the importance of a concept's being able to serve main cognitive and social purposes of the original from Mark Richard.

[37] The reader may worry that this argument presupposes a certain ontology of concepts, on which they have essential functions that cannot be eschewed. I do not intend anything this strong. I rely only on a background view that an important aim of our philosophical theorizing about concepts is to maintain contact with the phenomena to which everyday uses of our words refer. I draw inspiration from Bauer (2015), who points out that sometimes, as philosophers, "We do not feel a standing obligation to measure the distance between the range of everyday meanings of these words and the meanings we philosophers impose on them (Bauer 2015, p. 146)." We thus assume that we are making discoveries about real-world phenomena, when in fact we have changed the subject. For further explanation and defense of this view, see Bauer's (2015), especially chapter 8.

5.3.1 What's wrong with conservative evidentialism?

Stepping back, we can ask what epistemologists should want from a concept like *justification* in the first place. At bottom, we should want such a concept to evaluate the beliefs of human doxastic agents—and to evaluate those beliefs not in terms of how well they promote some ends external to themselves—but *as such*. Given, then, that humans are creatures who sometimes have significant control over our belief-forming processes, we should want a theory of justification that can take account of this.

In what sense do human doxastic agents sometimes have significant control over our belief-forming processes? We have the capacity to be actively involved *both* in the process of proportioning our belief to our evidence, *and* in the process of constituting the evidence itself. Take the first point first. Earlier, I pointed out that proportioning one's belief to the accessible evidence can require active cognitive effort. This is particularly true in complex and controversial cases, where the evidence often does not come in a tidy package such that the subject can simply observe it and adjust her belief accordingly. More colorfully: it is not as if an oracle reveals to the agent that there is a 94% chance that p is true, and the agent proportions her belief to the evidence by simply adopting a credence of .94. Rather, the proportioning can require effort, as in the case of accessing the defeat relation that the *MDC* refers to, or of scrutinizing one's evidence.

Secondly, proportioning one's belief to the evidence can also be a process of *constituting* one's evidence. This too is because in some cases, the type of information one finds out in the world may not wear the facts about what it supports on its sleeve. Elena thus has to do some cognitive work to figure out what the information she is given really amounts to vis-à-vis her belief. This work enacts the transition from what Kelly calls *narrow* to *broad* evidence. Again, the information Elena is given about the congenial and uncongenial studies is narrow evidence. Proportioning her belief to that evidence consists partly in making judgments about how credible it is and how much weight it merits. This includes, for instance, coming up with alternative hypotheses that could account for the data. In doing this, Elena's agency is actively involved in *constituting* the broad evidence to which she proportions her belief. In that sense, the process of proportioning her belief to the evidence is also a process of constituting the evidence. So, the subject's doxastic agency is not inert with respect to the proportioning process, *or* with respect to the evidence itself.

Evidentialism determines justification by evaluating whether the agent successfully proportioned her belief to her evidence. But in spelling out what this means, the conservative evidentialist leaves no room for taking stock of these ways in which the subject's agency is active in the *process*, both of doing the proportioning, and of constituting the evidence itself. The evaluative focus is exclusively on the end state of a belief's *being* proportioned to the evidence, and the rubric for assessing that state does not reflect the agent's role in producing it. If the purpose of our normative epistemic concepts is to evaluate the beliefs of beings whose agency is actively involved both in constituting their evidence and proportioning their beliefs to it, then this rubric is impoverished. Indeed, it would seem better suited to a world in which the role of epistemic agency in belief formation *is* limited to adopting whatever

credence the oracle recommends.[38] Metaphorically, then, the problem with the conservative evidentialist picture is that it removes the doxastic agent from the evaluative picture to too large a degree.

A more fitting model of justification would leave room to evaluate the agent's performance in constituting the evidence and proportioning her belief to it. Subjects *do better epistemically* when they do well in these aspects of the belief forming process. This is why, as argued in the previous section, we have reason to push one another to live up to our potential as doxastic agents in these ways, and to hold one another epistemically accountable when we fall short. This is not just a practical matter, but an epistemic one, so our normative epistemic concepts should leave room to account for it.

5.3.2 Liberal evidentialism?

There is in principle room within the evidentialist framework for a more *liberal* picture that spells out what it means to successfully proportion one's belief to her evidence in a way that takes account of how human agency can be involved in that process. But the resulting concept of evidentialist justification will not count polarized beliefs like Elena's as justified.

More specifically, there is in principle room within an evidentialist framework to give a theory of justification on which a subject's belief can be criticized for (i) and/or (ii). It would simply require giving a less restrictive answer to the hard question about accessibilism, thus raising the bar on what it means for an epistemic agent to successfully proportion her belief to the total accessible evidence. Giving up on (i) would mean that the agent is responsible not only for the evidence that she *in fact* has, but also for whatever evidence she *ought* to have as a result of meeting this standard.[39] Giving up on (ii) would allow evidentialism to set a standard for what it means to put adequate effort into taking stock of the different pieces of one's evidence, how they bear on each other, and what it amounts to as a whole. More broadly, it would mean that *accessible evidence* is a normative concept that includes whatever evidence a subject ought to have access to as a result of meeting a certain standard of exercising her doxastic agency in the belief forming process.[40]

[38] The conservative concept of justification may be better suited to evaluating simpler cases, like perceptual belief. Generally, considering one's perceptual evidence does not require intentional cognitive action on the part of the epistemic agent. Perhaps the conservative evidentialist had these cases in mind in constructing her theory of justification. But the model is not appropriately extended to these complex and controversial cases.

[39] The specific question of whether the evidence Elena would be normatively responsible for on this standard includes the evidence that she would have gotten from scrutinizing the congenial study is beyond the scope of my discussion, as I am focusing on evidence that she could now access simply by further reflecting on evidence already within her ken. But I do not think it is beyond the pale to suppose that her belief loses some measure of justification because she lacks this evidence. After all, Elena does not have any positive reason to think the congenial study will be free of flaws when she chooses not to scrutinize it.

[40] As suggested by my earlier empirical argument, this standard ought to be determined in part by facts about the subject's cognitive agency. For instance, competent adult human doxastic agents are fitting subjects for a higher standard of what it means to reflectively take stock of one's evidence than young chil-

This liberal notion of holding the agent responsible for her evidence in these senses is not in principle anti-evidentialist. Indeed, in some sense the conservative evidentialist *has* already embedded a standard of what we can expect from the doxastic agent in the definition of what it means to successfully proportion one's belief to her evidence. It is just a lower one. For instance, Kelly would presumably consider Elena's belief unjustified had she found the fatal flaw in the uncongenial study, remained aware of it as she revised her belief, but failed to treat it as a defeater of the support that E1 would otherwise lend to *Not-Deterrent*. This would count as a failure to successfully proportion her belief to her total evidence, since although she recognized the flaw, she failed to recognize the way it bears on E1, and to proportion her belief to her evidence in light of that.[41] Similarly, I imagine he would consider Elena's belief unjustified if she had found the fatal flaw in the congenial study, but then willed herself to forget that she had seen it, so her broad evidence continued to reflect that the congenial study's results support *Deterrent*.[42] To this extent, it does matter to the conservative evidentialist whether the subjects exhibits epistemic responsibility in the process of acquiring evidence and proportioning her belief to it. The difference between the conservative and liberal pictures lies simply in how they understand the range of our epistemic responsibility. This underlies their different answers to the hard question. The liberal claims that our capacities as doxastic agents render us fitting subjects for a broader range of responsibilities. Ultimately, a careful consideration of just how far this extends is beyond the scope of my discussion. Earlier, I argued that there is reason to think our ordinary concept of justification already holds us responsible for (i) and (ii), and that we should be wary of giving this up. And I have just argued that there is reason to think we are fitting subjects for this broader range of epistemic responsibility, given the capacities we have as doxastic agents, and how they are involved in belief formation. I leave the question of exactly how to delimit the range of our responsibility aside, but suggest that it be based on our best understanding of our capacities as doxastic agents, and the ways in which exercising them allows us to do well epistemically.

If what I have said is right, then Kelly must either adopt a more liberal definition of evidentialist justification, or concede that Elena's polarized belief is not justified

Footnote 40 (continued)
dren and nonhuman animals are, because the latter lack the type of cognitive agency I have been describing. Because the subject's agency is involved in constituting her broad evidence, and actively reflecting on what its overall weight and balance supports, she bears a normative epistemic responsibility for the evidence she ought to have as a result of doing these things well.

[41] This kind of case is the reason I said that Kelly's restrictive answer to the hard question makes justification *nearly* automatic, rather than absolutely so.

[42] It is unclear just *how* low the standard is on the conservative picture, but there is reason to think that it is quite low. Kelly does not take issue with Elena's judgment that the congenial study does not contain serious flaws, even though she has been presented with information that directly challenges this, in the form of criticisms of the congenial study. This might mean that an error in the opposite direction would also be deemed unproblematic—for instance, if the uncongenial study contained only a minor flaw, and Elena judged it to be a fatal one. This seems like the sort of thing that could problematically lead people to dismiss entire bodies of research.

on evidentialist grounds. Given what I said about the tenability of the conservative definition, I advise the former.

Acknowledgements I would like to thank Susanna Siegel, Mark Richard, Susanna Rinard, Zoe Jenkin, and Paul Blaschko for helpful comments on previous versions of this paper.

References

Alston, W. (1989). *Epistemic justification*. Ithica: Cornell University Press.

Batson, C. D. (1975). Rational processing or rationalization? The effect of disconfirming information on stated religious belief. *Journal of Personality and Social Psychology, 32,* 176–184.

Bauer, N. (2015). *How to do things with pornography*. Cambridge: Harvard University Press.

Baumeister, R. F., & Newman, L. S. (1994). Self-regulation of cognitive inference and decision processes. *Personality and Social Psychology Bulletin, 20,* 3–19.

Benoît, J.-P., & Dubra, J. (2019). Apparent bias: What does attitude polarization show? *International Economic Review, 60*(4), 1675–1703.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39,* 752–766.

Chaiken, S., Giner-Sorolla, R., & Chen, S. (1996). Beyond accuracy: Defense and impression motives in heuristic and systematic information processing. In P. Gollwitzer & J. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 553–578). New York: Guilford.

Chen, H. C., Reardon, R., Rea, C., & Moore, D. J. (1992). Forewarning of content and involvement: Consequences for persuasion and resistance to persuasion. *Journal of Experimental Social Psychology, 28,* 523–541.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 73–98). New York: Guilford.

Clark, J. K., Wegener, D. T., & Fabrigar, L. R. (2008). Attitudinal ambivalence and message-based persuasion: Motivated processing of proattitudinal information and avoidance of counterattitudinal information. *Personality and Social Psychology Bulletin, 34,* 565–577.

Clark, J. K., Wegenes, D. T., Habashi, M. M., & Evans, A. T. (2012). Source expertise and persuasion: The effects of perceived opposition or support on message scrutiny. *Personality and Social Psychology Bulletin, 38,* 90–100.

Conee, E., & Feldman, R. (2004a). *Evidentialism: Essays in epistemology*. Oxford: Oxford University Press.

Conee, E., & Feldman, R. (2004b). Evidence. In Q. Smith (Ed.), *Epistemology, new essays*. Oxford: Oxford University Press.

Conee, E., & Feldman, R. (2008). Evidence. In Q. Smith (Ed.), *Epistemology: New essays*.

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology, 63,* 568–584.

Ditto, P., Scepansky, J., Munro, G., Apanovitch, A., & Lackhart, L. (1998). Motivated sensitivity to preference-inconsistent information. *Journal of Personality and Social Psychology, 75,* 53–69.

Dogramaci, S. (2011). Reverse engineering epistemic evaluations. *Philosophy and Phenomenological Research, 84,* 513–530.

Dogramaci, S. (2015). Communist conventions for deductive reasoning. *Noûs, 49,* 776–799.

Earman, J. (1991). *Bayes or bust? A critical examination of Bayesian confirmation theory*. Cambridge: MIT Press.

Fitelson, B. (1999). The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science (Proceedings Supplement), 66,* 362–378.

Freund, T., Kruglanski, A. W., & Shpitzajen, A. (1985). The freezing and unfreezing of impressional primacy: Effects of the need for structure and the fear of invalidity. *Personality and Social Psychology Bulletin, 11,* 479–487.

Gawronski, B., & Bodenhausen, G. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132,* 692–731.

Gerber, A., & Green, D. (1999). Misperceptions about perceptual bias. *Annual Review of Political Science, 2,* 189–210.

Giner-Sorolla, R., & Chaiken, S. (1997). Selective use of heuristic and systematic processing under defense motivation. *Personality and Social Psychology Bulletin, 23,* 84–97.

Ginossar, Z., & Trope, Y. (1987). Problem solving in judgment under uncertainty. *Journal of Personality and Social Psychology, 52,* 464–474.

Glaeser, E., & Sunstein, C. (2013). *Why does balanced news produce unbalanced views?* NBER working paper series paper no. 18975. Retrieved July 22, 2019, from https://www.nber.org/papers/w18975.

Harmon-Jones, E., & Harmon-Jones, C. (2008). Cognitive dissonance theory: An update with a focus on the action-based model. In J. Y. Shah & W. L. Gardner (Eds.), *Handbook of motivation science* (pp. 71–83). New York: Guildord Press.

Hart, W., Albarrcín, D., Eagly, A., Brechan, I., Lindberg, M., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin, 135,* 555–588.

Hastrof, A. H., & Cantril, H. (1954). They saw a game. *Journal of Abnormal Social Psychology, 49,* 129–134.

Houston, D. A., & Fazio, R. H. (1989). Biased processing as a function of attitude accessibility: Making objective judgments subjectively. *Social Cognition, 7,* 51–66.

Jern, A., Chang, K., & Kemp, C. (2009). Bayesian belief polarization. In *Proceedings of the 22nd international conference on neural information processing systems (NIPS), Vancouver, Canada, 7–10 December 2009*, viewed 8 September, 2019. https://papers.nips.cc/paper/3725-bayesian-belief-polarization.

Jern, A., Chang, K., & Kemp, C. (2014). Belief polarization is not always irrational. *Psychological Review, 121*(2), 206–224.

Jonas, E., & Frey, D. (2003). Information search and presentation in advisor–client interactions. *Organizational Behavior and Human Decision Processes, 91,* 154–168.

Kelly, T. (2003). Epistemic rationality as instrumental rationality: A critique. *Philosophy and Phenomenological Research, 66,* 612–640.

Kelly, T. (2008a). Disagreement, dogmatism, and belief polarization. *The Journal of Philosophy, 105,* 611–633.

Kelly, T. (2008b). Common sense as evidence: Against revisionary ontology and skepticism. *Midwest Studies in Philosophy, 32,* 53–78.

Kelly, T. (2008c). Evidence: Fundamental concepts and the phenomenal conception. *Philosophy Compass, 3,* 933–955.

Kluegel, J. R., & Smith, E. R. (1986). *Beliefs about inequality: Americans' views of what is and what ought to be*. New York: De Gruyter.

Koehler, J. J. (1993). The influence of prior beliefs on scientific judgment of evidence quality. *Organizational Behavior and Human Decision Processes, 56,* 28–55.

Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology, 53,* 636–647.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108,* 480–498.

Kunda, Z. (1999). *Social cognition: Making sense of people*. Cambridge, MA: MIT Press.

Lackey, J., & Sosa, E. (2006). *The epistemology of testimony*. Oxford: Oxford University Press.

Leippe, M. R., & Elkin, R. A. (1987). When motives clash: Issue involvement and response involvement as determinant of persuasion. *Journal of Personality and Social Psychology, 52,* 269–278.

Liberman, A., & Chaiken, S. (1992). Defensive processing of personally relevant health messages. *Personality and Social Psychology Bulletin, 18,* 669–679.

Liu, C.-H. (2017). Evaluating arguments during instigations of defence motivation and accuracy motivation. *British Journal of Psychology, 108,* 296–317.

Lord, C., Ross, L., & Lepper, M. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37,* 2098–2109.

Lundgren, S. R., & Prislin, R. (1998). Motivated cognitive processing and attitude change. *Personality and Social Psychology Bulletin, 24,* 715–726.

Miller, A. G., McHoskey, J. W., Bane, C. M., & Dowd, T. G. (1993). The attitude polarization phenomenon: The role of response measure, attitude extremity, and behavioral consequences of reported attitude change. *Journal of Personality and Social Psychology, 65,* 561–574.

Murno, G. D., & Ditto, P. H. (1997). Biased assimilation, attitude polarization, and affect in reactions to stereotype-relevant scientific information. *Personality and Social Psychology Bulletin, 23,* 636–653.

Petty, R. E., & Wegener, D. T. (1999). The elaboration likelihood model: Current status and controversies. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 41–72). New York, NY: Guilford Press.

Pittman, T. S., & D'Agnostino, P. R. (1985). Motivation and attribution: The effects of control deprivation on subsequent information processing. In G. Weary & J. Harvey (Eds.), *Attribution: Basic issues and applications* (pp. 117–141). New York: Academic Press.

Pollock, J. (1986). *Contemporary theories of knowledge*. Towota, NJ: Rowman and Littlefield Publishers.

Pomerantz, E. M., Chaiken, S., & Tordesillas, R. S. (1995). Attitude strength and resistance processes. *Journal of Personality and Social Psychology, 69,* 408–419.

Pyszczynski, T., & Greenberg, J. (1987). Toward and integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 297–340). New York: Academic Press.

Schuette, R. A., & Fazio, R. H. (1995). Attitude accessibility and motivation as determinants of biased processing: A test of the MODE model. *Personality and Social Psychology Bulletin, 21,* 704–710.

Srull, T. K., & Wyer, R. S. (1986). The role of chronic and temporary goals in social information processing. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition* (pp. 503–549). New York: Guilford Press.

Taber, C., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *Journal of Political Science, 50,* 755–769.

Tetlock, P. E. (1985). Accountability: A social check on the fundamental attribution error. *Social Psychology Quarterly, 48,* 227–236.

Tetlock, P. E., & Kim, J. I. (1987). Accountability and judgment processes in a personality prediction task. *Journal of Personality and Social Psychology, 52,* 700–709.