Defeaters and Disqualifiers

Daniel Muñoz

MIT

Forthcoming in *MIND*.

Penultimate draft of July 25, 2018

Abstract

Justification depends on context: even if E on its own justifies H, still it might fail to justify in the context of D. This sort of effect, epistemologists think, is due to the possibility of defeaters, which undermine or rebut a would-be justifier. I argue that there is another fundamental sort of contextual effect, disqualification, which doesn't involve rebuttal or undercutting, and which cannot be reduced to any notion of screening-off. A disqualifier makes some would-be justifier otiose, as direct testimony sometimes does to distal testimony, and as manifestly decisive evidence might do to weak but gratuitous evidence on the same team. Basing a belief on disqualified evidence, moreover, is irrational in a distinctive way. One is not necessarily irresponsible. Instead one is turning down, for no reason, an upgrade to a sleeker, stabler basis for one's beliefs. Such an upgrade would prevent wastes of epistemic effort, since someone who bases her belief on a disqualified proposition E will need to remember E and rethink her belief should she come across a defeater for E. The upgrade might also reduce reliance on unwieldy evidence, if E is relevant only thanks to some labyrinthine argument; and to the extent that even ideal agents should doubt their ability to follow such an argument, even they should care about disqualifiers.

Introduction

Not to be a gossip, but I just heard Alice say that Bob said *p*. Now, Bob is a reliable source when it comes to matters *p*-related—and ditto for Alice on things Bob-related—so Alice's testimony justifies my belief that *p*. But her testimony isn't *guaranteed* to be a justifier. Contrast two cases:

Just Alice All I've got to go on is Alice's testimony; it justifies belief in *p*.

Bob, Too Alice shows me the letter she got from Bob. Sure enough, it says *p*.

In the second case—suitably tightened—Bob's testimony alone justifies. Alice's testimony is there, but Bob's word *disqualifies* it from giving any justification, making it unnecessary or gratuitous. That is why, with both bits of testimony in hand, it would be irrational for me to base my *p*-belief on Alice instead of going closer to the source.

Disqualifiers take a would-be justifier and make it irrelevant. So they are like *defeaters,* which can make a would-be justifier irrelevant by ruining the case it makes for a hypothesis, either by *undercutting* the link between evidence and hypothesis, or *rebutting* the hypothesis directly (Pollock & Cruz 1999, pp. 196–7). But notice that Bob's testimony does no such thing. If anything, hearing from Bob has the opposite effect: it confirms *p* and vindicates Alice. Disqualifiers don't defeat. Instead of ruining a justifier, they *obviate* it.

Disqualification is also like *screening-off*. Bob's word screens Alice's in the sense that, given that Bob said *p*, the presence of Alice's testimony won't boost or hurt my justification for *p*. But disqualification is a finer notion than screening: two facts may probabilistically screen each other, though only one disqualifies the other, and some evidential screens work just by defeating. So we need disqualifiers, and we can't reduce them to defeaters or screens. That's the claim, anyway. To defend it, we need to start from the top and go a lot more slowly.

1. Disqualifiers

I heard Alice say that Bob wrote to her saying *p*. Alice then showed me the letter:

> Dearest Alice,
> *p*.
> Yours forever,
> Bob.

Bob's testimony (that *p*) *disqualifies* Alice's testimony (that Bob said *p*) from justifying my belief in *p*: once I have read the letter, my justification comes from Bob and Bob alone.[1] Alice's word is now irrelevant—otiose, idle—and if it remains part of the basis for my belief that *p*, there is something irrational about me. I am looking an epistemic upgrade in the face and saying, 'No, thank you', and for no intelligible reason.[2]

What's the nature of the upgrade? Here's the rough idea (polished in §5). Before I see the letter, I have to base my belief on two pieces of testimony: Alice's and Bob's. Why Alice's? Because it is my only basic reason for believing *p*. Why Bob's? Because it is the only link between what Alice says and whether *p*. So I need both: justification for *p* has to flow from Alice, through Bob. But once I see the letter, I don't need Alice anymore. I can get away with basing just on Bob—which would relieve me of certain responsibilities, like remembering what Alice said, and reopening the question of whether *p* should I ever learn she can't be trusted.

[1] To keep things simple, I assume that justifiers, defeaters, and disqualifiers are *propositions*, like the proposition that *p*. Another view is that they are all propositional *attitudes*, like my belief that *p*. This difference won't matter for my main examples, where the disqualifiers are all known to be true.

[2] What happens if I only have Alice to go on? Then the fact that I haven't seen the letter *qualifies* Alice's testimony to be my justification: the lack of superseding evidence explains why Alice's testimony is needed, and therefore relevant. A more general example of a qualifier is that I *at most* have such-and-such evidence. For background, see Yablo 2012, Skow 2016, pp. 80–1, 108–9, 173n.48, and Baron-Schmitt ms. on the distinction between 'ennoblers' (qualifiers) and 'enablers' (the opposite of undercutting defeaters).

Bob's letter disqualifies Alice's testimony because the testimony bears on *p* only in so far as it bears on the letter. The letter's evidential contribution, in this sense, 'subsumes' the testimony's. But not all disqualifiers are subsumers. *Overwhelmers* are more like glory-hogs:

Solo	The wall looks red to you. Lo: justification for belief in its redness.
Guru	The guru, you are rationally certain, cannot but speak the truth. She tells you that the wall is red. You look at it, and red it looks.

If you are *really* rationally certain that the guru speaks the truth, and just as rightly sure that the guru has pronounced the wall to be red, then it doesn't matter that the wall looks red to you. You are justified no matter how it looks. Just think: what would have happened had the wall looked *green*? You would, quite reasonably, have concluded that you were being visually tricked. And what if you hadn't seen the wall? You should have still been certain of its redness.

The guru's declaration thus disqualifies your experience from justifying your belief that the wall is red: the declaration leaves no *room* for the experience to play any role. The experience is gratuitous, otiose, irrelevant. But it is not subsumed, in the way that Bob's letter subsumes Alice's testimony. It is not as though your red-experience is relevant *only in so far as* it bears on the guru's pronouncement. Rather, in general, evidence is relevant only when there is nothing on the scene so awesome that it trumps all nearby would-be justifiers. The guru's word, I am imagining, would trump your experience, but not by subsuming its evidential upshot.[3]

In sum: *disqualified evidence is irrelevant because it's unnecessary*. To count the testimony

[3] Subsumers, in general, grant us shorter chains of justification. But not everything that brings short chains will count as a subsumer. Suppose I've heard from two people: Alice says 'Bob said *p*', and Carol says '*p*'. Thanks to Carol, I have a shorter path to *p* than I could have gotten with just Alice. But here's the test: does Alice's word bear on *p* only in so far as it bears on what Carol said? No. The shorter chain won't always subsume. (Thanks to an anonymous referee whose comments helped me see this.)

with the letter is to count double; to count your vision with the guru is to count without need.

Objection: these cases are under-described. Sticking to the Alice-Bob Chronicle for now, there are plenty of natural ways of fleshing out the story so that Alice's testimony will end up providing justification—maybe only *some* justification, but good justification regardless. What if I'm not certain that I've seen Bob's letter aright? What if Alice is herself an expert vis-à-vis whether *p*, and I know that she would not mislead me by transmitting faulty testimony?

Response: sure, there are lots of nearby cases where Alice's testimony bears on *p*. Hearing from the source won't always disqualify the transmitters. But in this sort of case, whenever there *isn't* disqualification, there must be an explanation. And once we identify the typical explanations we can stipulate that they aren't at work in our case, making it a suitably tight exemplar of disqualifying.

Imagine a chain of testifiers, some 'closer' and some more 'distant', each of which bears on a hypothesis. The closest link is the one that bears on the hypothesis directly. (Zed said that he has a headache—which bears on the headache.) The next, more distant link bears directly on the one before it. (Yuna said that Zed said that he has a headache.) And so on for the rest of the links. (Alice said that Bob said that…Yuna said that Zed said that he has a headache.)[4]

Now, in general, closer testimony will disqualify more distant testimony. But closer testimony won't disqualify distant testimony that is *critical*, in the sense of having some independent probative value besides serving as an indicator of the closer testimony. Consider:

[4] Note that we are *assuming* that the distant testimony (from Alice–Yuna) could unconditionally support the hypothesis (that Zed has a headache). Why is this assumption necessary? Doesn't it just follow from the fact that each testifier confirms the next? No: confirmation isn't necessarily transitive. Consider a case from Fitelson 2012: E1 says that I drew a black card; H1 says that I drew a black ace; H2 says that I drew an ace. E1 confirms H1, which confirms H2, but E1 doesn't confirm H2.

The Times & the AP I read in the *Times* that the AP says that *q*. I then read the AP wire, to the effect that *q*. But I take it that if the *Times* printed it, they probably have vetted the story in some way that I haven't.

In this case, I should take the *Times'* report to corroborate the AP. There is a real chance, I think, that had the AP report been false or unreliable, the *Times* would have refused on that basis to pass it along to me. So my reading the AP wire doesn't wholly disqualify the *Times'* testimony from being relevant: the value of the *Times* outstrips its use as an indicator of what the AP says.

There is another way in which distant testimony can be helpful:

Bob, Too 2 I discover that Bob's letter says *p*, but I'm not certain that I've read it right. Alice's testimony might still be helpful here.

Maybe I am not sure that I understand Bob's sense of humor. (Is he saying *p* sarcastically?) Or maybe my vision is worse than my hearing. Or maybe I only saw the '*p*' part of the missive, and I think that there is some possibility that I was looking at an upside-down '*d*'. In any such case, Alice's testimony will still be a reason to believe *p*, and will still partly ground my justification.

But you get the point. Assuming that Alice's testimony is *merely* indicative of Bob's, and that I have an impeccable independent grip on what Bob said, possessing Bob's testimony will make Alice's redundant—and the redundancy here will be asymmetric in Bob's favor.[5] Assuming that I know the *Times* to be *merely* transmitting the AP report, the report and transmission will be redundant, and only the report will be relevant.

[5] Similar things will hold of others chains whose distal links bear on the hypothesis only via the closer links. Instead of testimony, the chain links be facts about directly causally relevant events; see e.g. Reichenbach 1956, p. 189. Or consider these bits of evidence for Zed's being sad: I hear Zed sobbing; I hear the echo of Zed's sobbing; I hear the echo's echo; etc. For more cases of subsumption, see Barnett (forthcoming, §6.2) on dependent belief; Goldman's (2001, p. 99) guru; Kelly's (2005, fn. 20) imperfect transcript; Lackey's (2014, p. 249) treatment of non-autonomous complete source-dependence, and *Two Instruments* in §3, below.

(A quick aside. What would symmetric redundancy be like? Something like this:

Print & Web The Texas Tribune reports on the AP wire (to the effect that *q*) both in print and online. (Someone types the report into a computer, and this type-up is immediately fed into both media.) The print and web are redundant, but neither supersedes the other.

Alice & Alicia One time, Alice and Alicia both heard Bob say *p*. I hear both of them testify as much. But since I know Alice and Alicia are honest epistemic equals, the testimony of one adds nothing to the testimony of the other, though neither disqualifies the other.

I am rationally permitted to base my belief that *p* on *either* Alice or Alicia's testimony, and I may believe *q* on the basis of either the print or web versions of *The Tribune*. But I may not double-count. When building a case for *p* I may count either Alice or Alicia, but not both at once; their testimony isn't additive.[6] And this will be true regardless of whether Alice and Alicia are critical testifiers, so long as neither is critical in a way that outstrips the other.)

Now back to the Alice-Bob-*p* primal scene. How can we make sure that it's a case of disqualification, as advertised? We just need to stipulate that Alice's testimony is *unnecessary* and *uncritical*. There is no way in which Alice's word bears on *p* except by indicating what Bob said, and I can know what Bob said without Alice's help.

What about the guru and the vision? Surely my experience would be relevant to some extent if I needed confirmation in the guru's powers, or if her oracular pronouncements left room for complementary justification—but I don't, and they don't. *Given* that I have impeccable

[6] For discussions of non-additivity in ethics, see Kamm 1983, Kagan 1988, Dancy 2004, and Johnson King forthcoming. Back in epistemology, Kotzen (ms., §6) gives an example like *Print & Web*, holding that 'intuitively' the one source is 'irrelevant' if I've already encountered the other, and that this irrelevance is due to redundancy rather than undercutting defeat. But why should it matter which one I see first? Kotzen is right that the case involves some redundancy, but I think he treats the redundancy as asymmetric when it really isn't. (For other epistemic examples, see Muñoz ms.)

access to the guru's word *and* full knowledge that she speaks the truth, there is no need for my vision to justify my belief that the wall is red.[7]

Of course, in real life, there are no gurus—people whose testimony is known to be perfect. There are hardly even any Bobs—people whose testimony is perfectly known. Our access to contingent worldly facts is imperfect, which means that supporting evidence—Alice's testimony, my vision of the wall—tends to be at least somewhat useful and relevant. But even so, this supporting evidence can still be made *less* relevant by virtue of being *more* otiose than it would have been. Disqualification can come in degrees. After I've seen Bob's letter, Alice's testimony may still give me *some* justification for *p*, since I can't absolutely trust my sight; still, her testimony gives less justification than it did before. After I've heard from the guru, I may still get some support from my own vision, since I can't absolutely trust my choice of oracle; still, the guru's proclamation mostly supersedes what I get from my eyesight.

So overwhelming and subsuming aren't always all-or-nothing. When they are, as in the idealized Bob and guru cases, we are dealing with *total* disqualifiers, which make things wholly irrelevant. In real-world cases, we find only *partial* disqualifiers (or 'depreciators'), which just make things less probative to some degree. Total disqualification is cleaner and more striking, which is why we looked at it first, but partial disqualification is equally real and undoubtedly more common.

[7] One last wrinkle. Perhaps, even though I know that the guru speaks the truth, and on that basis infer that contra-guru evidence is misleading, my trust in the guru could be undermined by what I later experience. (This kind of case comes up in discussions of Kripke's 'dogmatist paradox'; see Harman 1973, pp. 148–9; Kripke 2011; Lasonen-Aarnio 2014a; see also Elga 2007, p. 483 and Zagzebski 2012, p. 116.) A potential example: the guru says that my chances of having a red experience are a measly $1*10^{-999}$ —then I have one. But never mind this. We can stipulate that nothing so special is happening in our case.

2. Defeaters

You might think that ‘disqualifiers’ for some evidence E to bear on a hypothesis H are really a familiar thing—*defeaters*. For defeaters and disqualifiers have something key in common: both explain how E may justify H in one scenario but not elsewhere. Here is Kelly (2016, §1, emphasis original):

> Even if evidence *E* is sufficient to justify believing hypothesis *H* when considered in isolation, it does not follow that one who possesses evidence *E* is justified in believing *H* on its basis. For one might possess some additional evidence *E′*, such that one is not justified in believing *H* given *E* and *E′*. In these circumstances, evidence *E′ defeats* the justification for believing *H* that would be afforded by *E* in its absence. Thus, even if I am initially justified in believing that *your name is Fritz* on the basis of your testimony to that effect, the subsequent acquisition of evidence which suggests that you are a pathological liar tends to render this same belief unjustified.

Thus the job of a defeater D is to ruin someone’s justification for H, which they would have gotten from E on its own. The typical drama: a thinker has justification thanks to E, but then learns D, and *poof* goes the justification.

But already this is unlike the case of Alice and Bob. When I get Bob’s letter, I don’t lose justification—if anything I end up *gaining*. The only thing lost is my reliance on Alice, whose testimony becomes gratuitous as evidence for *p*. The difference between disqualifiers and defeaters is even starker when we consider the varieties of defeat. Defeaters come in two kinds:

Rebutting Defeat D defeats E by providing evidence against H.

Undercutting Defeat D defeats E by undermining the E-H connection.[8]

But Bob’s letter doesn’t rebut *p*: it confirms *p*. And it doesn’t undermine Alice: it vindicates her.

[8] This distinction began with Pollock’s (1970, 1974) treatment of Type I and Type II defeaters. Now the standard terms are ‘rebutting’ and ‘undercutting’; see Pollock and Cruz 1999, pp. 196–7, Kelly 2016, and Sudduth (no date). For an early paper on defeat, see Lehrer and Paxson 1969. For a more recent addition to the literature—‘higher-order’ defeaters—see Christensen 2008 and Lasonen-Aarnio 2014b.

(What would a real defeater look like? Alice is undercut if I see that Bob's letter doesn't say *p*; my vision is rebutted if the guru says 'The wall isn't red!'. Note that these defeaters don't just fail to *add up* with Alice and the red-experience: instead, they *subtract* from my justification. And notice also a parallel between subspecies: rebutters and overwhelmers both target a hypothesis directly; undercutters and subsumers both problematize a path from the evidence.)

There is one more difference between defeaters and disqualifiers. If my evidence is defeated, and I continue to believe on the basis of it, there is something *epistemically irresponsible* about me. (The phrase is from Schechter 2013, p. 432.) I am according the evidence more weight than it really has, treating it as more of an indicator than it really is. As a result, I am liable to end up having baseless, unjustified beliefs—an epistemic Wile E. Coyote, stepping off the cliff without a plank to stand on. But this is not what it is like to base beliefs on disqualified evidence. Suppose I've heard from Alice and Bob, and I believe *p* on the basis of both instead of Bob alone. Is my belief unjustified? Am I *irresponsible*? If I am, it's hard to say why. Clearly, Bob's word is a fine basis. And Alice's, which would have been fine on its own, hasn't been defeated. What could be irresponsible about basing on both?

The problem here is not that of Wile E. Coyote walking baseless over the cliff. It is rather that of someone who only needs to rely on one thing—but then relies on two. If I keep leaning on Alice's word after getting hold of Bob's, there is something wrong with me, even given that Alice's testimony would have been fine on its own, and even given that there are no defeaters in sight. I'm not irresponsible, but I am *overly attached*—clinging to evidence that I don't need, making no effort to prune my bases. Because Alice is uncritical, she cannot add to my justification, given that I already know full well that Bob said *p*. But in basing my *p*-belief on

Alice, even partly, I treat her like a real contributor. This is a new kind of mistake, orthogonal to irresponsibility.[9]

Bottom line: Bob's testimony, a disqualifier, doesn't work like a defeater. Disqualification isn't defeat. We need another notion.

3. Probabilistic screening-off[10]

Defeaters aside, have epistemologists talked about anything else that works like a disqualifier? The closest concept is due to Weatherson (ms.), who discusses *evidential screening-off*. But first, we should look at a more austere notion familiar from Bayesian epistemology: *probabilistic* screening-off.[11]

What is it for S to probabilistically screen off some would-be evidence E from H? Two conditions need to hold:

(1) E probabilistically supports H—i.e. $\Pr(H \mid E) > \Pr(H)$.

(2) Given S, E and H are probabilistically independent—i.e. $\Pr(H \mid E \wedge S) = \Pr(H \mid S)$.

Simple enough. This notion of screening is important, crisp, and impressively well-understood.[12] There are some general objections to the reduction of evidential support to

[9] A bonus fact about disqualifiers: one can't know that *p* if one's *p*-belief is based on defeated evidence, but (it seems) one can know that *p* even if one's belief in *p* is partly based on *disqualified* evidence. If that's right, knowledge is consistent with (a kind of) irrational basing.

[10] My thanks to an anonymous referee for detailed, invaluable advice about §§3 and 4. The arguments are now clearer, and I hope that §4 is now more fair to Weatherson.

[11] Also relevant is Zagzebski's (2012, pp. 102–7) notion of a 'preemptor'—an epistemic reason that 'replaces' one's old reasons. (For example, an authority's judgement might replace my private evidence.) This is an interesting concept, but it is unclear how 'replacement' relates to undercutting and subsuming. So I take it that preemption by itself can't be used to distinguish defeaters from disqualifiers.

[12] See Reichenbach 1956; Atkinson and Peijnenburg 2013. Note that other writers use other definitions of 'screening off'. For some, it is enough that E and H are independent conditional on S. For

probabilistic support. But set them aside. We are interested in another question.

Can we reduce disqualifiers to probabilistic screens? The simplest reduction would identify screens with disqualifiers: to disqualify just is to screen. Easy enough, but there are also easy counterexamples: junky screens. Let E be one's evidence for H, and let S be some irrelevant contingent proposition (S is 'junk', for the purposes of justifying H). E ∧ S will screen E from H. But by hypothesis, E is relevant to H, since it's one's evidence. The conjunction screens its conjuncts, but it does not make them irrelevant as a disqualifier would; so, not all screens disqualify. This seems right, but not too deep. The reduction just needs another condition: disqualifiers are probabilistic screens such that no logically weaker proposition is also a screen.

A deeper kind of counterexample has to do with cases where S *does* make E irrelevant, but not by disqualifying. Just saying 'S screens E from H' leaves open how S makes E and H independent: S might subsume or overwhelm E, but it could also undercut E or rebut H decisively. So not all screens disqualify; some are just defeaters. Again, seems right, but non-lethal. Defeaters leave distinctive ripples in a space of probabilities; for example, if S screens E by decisively rebutting H, then Pr(H|S) = 0. This suggests another way to refine the reduction: disqualifiers are screens that don't give off the telltale ripples of defeat, whatever those may be.

But there is a final, fatal problem for *any* reduction to screens, no matter how refined:

Titelbaum (ms., p. 63), E must also unconditionally depend on H. For Shogenji (2003, p. 614), unconditional dependence isn't needed, but E and H must be independent conditional on S *and* on ¬S. Nothing hangs on this disagreement. But to avoid confusion, (i) I define 'screening off' only over cases where E unconditionally supports H, since it is only these cases that I discuss; and (ii) I don't require independence conditional on ¬S, since that would allow for boring cases of disqualifying without screening. (<The guru said *q*> disqualifies <My experience suggests *q*> from being evidence for <*q*>, but my experience is good evidence conditional on the guru's *not* saying *q*. This isn't a deep problem for reducing disqualifiers to screens.)

two cases can be probabilistically alike but different in disqualification. What is relevant to what doesn't supervene on the relevant probabilities.

Consider a case with the same Alice-Bob structure as before. I'm standing in front of a curtain, behind which lie two digital thermometers. I don't know whether the second is giving out a reading, but I do know how the two are set up:

Two Instruments T1 tracks the weather; T2 tracks T1.

As background, assume (i) that I have no independent evidence about the weather, or about deviant ways in which my thermometers might indicate the weather; (ii) that I am rationally certain that T1's reading accurately tracks the weather; and (iii) that I am also rationally certain that T2's reading—if it gives one—will be an accurate report on what T1 has said; if T2 reads '*x*°', that is because T1 reads '*x*°'—and for no other reason. (T2 is an 'uncritical testifier'.) Against this background, the probability that the weather is 100° is 1, conditional on either thermometer's saying so.

Now suppose that I have only T2 revealed to me; it says '100°'; and so I am directly justified in believing that it says '100°'. Curious about our friend T1, I reckon that it too must read 100°, and on the basis of T1's reading, I infer in turn that the weather is 100°. Now, notice that <T1 reads 100°> and <T2 reads 100°> screen each other from <The weather is 100°>, but neither disqualifies the other.[13] T1's reading *has* to be relevant: T2 bears on the weather only in so far as it tells me about T1! But T2's *also* has to be relevant—it's my only access to T1! And yet

[13] '<*p*>' denotes the proposition that *p*, by the way. And when I say that these propositions screen each other, I mean that they are screens relative to the background given by assumptions (i)–(iii). Relative screening is easily defined: S screens E from H relative to B *iff* (1) given just B, E supports H; and (2) given S ∧ B, E and H are independent. (Absolute screening can be defined as the special case where B is trivial.)

the readings screen each other, given my background evidence, because whether I have one or both, the probability that the weather is 100° remains the same—exactly 1.

So far, nothing too new. Just another pair of screens that don’t disqualify. Here’s the twist. Without changing any relevant *probabilities*, we can introduce a disqualifier. Consider:

Two Instruments 2 T1 and T2 are set up as before. I see both; they read ‘100°’. Seeing T1 changes how I’m justified: I no longer need T2 to be certain in <T1 reads 100°>, because my experience justifies me directly. Now T1’s reading disqualifies <T2 reads 100°> from justifying <The weather is 100°>. But the probabilities involved are still the same—exactly 1. From a probabilistic point of view, there is no relevant difference between my current state and my state from before, when I could only see T2. And yet only the current case involves a disqualifier. If that’s right, then we have a counterexample not just to more-or-less refined reductions of disqualifiers to screens, but to *any* probabilistic reduction.

So we have a pair of problematic cases. What makes them work? The key feature is that T2—in addition to being ‘uncritical’—is an *incorrigible indicator* of T1: whatever T2 tells us about T1 is certain to be correct. For this reason, also seeing T1 doesn’t make it more probable that the weather is 100°, or even that T1 reads 100°. (Usually, epistemic middlemen are fallible, and hearing from the source will boost one’s confidence.) Because the distal evidence (T2) is incorrigible, it screens the proximal evidence (T1); and because the distal evidence is *uncritical*, it is disqualified given direct access to the proximal evidence. Once you see how the cases work, it’s not hard to think of more. But one pair is enough for now.

4. Evidential screening-off

Disqualifiers can't be reduced to probabilistic screens: probabilistic support relations are coarse, and some disqualifiers leave no trace among our rational credences. But there is a finer-grained, non-probabilistic notion of screening—*evidential screening off*.

Intuitively, for S to evidentially screen off E from H is for S to 'subsume whatever evidential force' E has (Weatherson ms., p. 1). (Note that Weatherson does not have in mind my distinction between 'subsuming' and 'overwhelming'—see §1.) More technically, for S to screen off E from H is for the following two conditions to hold:

(1) E is evidence for H.

(2) E ∧ S is no better evidence for H than S is, and ¬E ∧ S is no worse evidence for H than S is. (Weatherson ms., p. 2)

Weatherson also gives some examples, which are interestingly different from mine.[14]

Could we reduce disqualifiers to evidential screens? The reduction has some hope. Since evidential screening involves more than probabilities, the thermometer cases aren't a fatal problem; and thanks to (2), there can't be 'junky' evidential screens (¬E ∧ S *will* be worse evidence for H, if S is irrelevant).

But there is still the problem of coarseness. As before, there are many ways for a screen to do its work: S could screen E from H by subsuming E; S could support (or rebut) H so decisively that E no longer has a chance of justifying it; or S could undercut the link between E

[14] One stylized case involves subsuming (the suspect's fingerprints are screened off as evidence for his guilt by the fact that he was at the scene); one real-world case involves voting statistics (*given* that one is pro-choice, being from Massachusetts no longer predicts voting Democrat); see Weatherson ms., p. 2. Also notable: when two pieces of evidence are symmetrically redundant, they often screen each other (see *Alice & Alicia* and *Print in Web* in §1). But it's hard to classify these cases further. Do the two facts partially disqualify each other? Or is their conjunction somewhat undercut?

and H, while supplying some independent evidence for or against H. Some screens are defeaters; only some disqualify.

Weatherson's notion is good and important, but it doesn't automatically yield a reduction of disqualification. We could try forcing one. But for now, I think the best way to learn more about disqualifiers isn't by reducing them to something else, but by taking them as primitive and asking more exploratory questions. What makes disqualifiers unique? What do we lose by ignoring them? More bluntly, why care?

5. The rationale for disqualification[15]

Disqualifiers are intuitive and irreducible. Intuitive, because we can see them at work in cases—Alice and Bob, the wall and the guru, T1 and T2. Irreducible, because the cases feature more than what's already in the epistemologist's menagerie—rebutting and undercutting defeaters, evidential and probabilistic screens. Disqualifiers are a real addition, and they seem to make good sense.

But where do they come from? Even if we do intuitively take disqualified evidence to be irrelevant, is there any rationale behind our intuition? Is there some deeper fact about us, as epistemic agents, that explains why we should care about disqualifiers?

The deep fact, I think, is that we have to make do with our cognitive limits. Of course, we can imagine *unlimited* reasoners with endless time, storage, and cognitive horsepower, who

[15] This section was greatly improved by several rounds of comments from a second referee, who suggested the third part of the rationale, as well as its relation to self-doubting ideal agents; my sincerest thanks for the help, and for a very nice phrase—the 'epistemologist's menagerie'.

would have no trouble updating their worldview in response to any new information (e.g. by conditionalizing on total evidence), and who would have no real interest in streamlining this process. But we ourselves are keen to keep things simple. We are cruelly short on space and time, and we have awfully finite powers of inference and memory.

Now, limits or no limits, everyone has to worry about having unjustified beliefs. But only finite thinkers are hurt by *inefficient management*. Because of limits on storage space, we can't keep track of every fact that might conceivably become relevant. Because of limits on time, we can't update every one of our beliefs whenever we learn something new. So, for us, any policy that simplifies updates (without making us dumber) comes as epistemic manna.

Consider a mundane sort of update: I am in my room, inquiring into whether *q*, when a coconut falls onto my head. I ought to review *some* of my opinions—forming the belief that there's a coconut nearby, dislodging the belief that I am safe from falling objects, reconsidering whether I'm alone in here. But I shouldn't bother (just for the occasion) reopening questions about what makes right acts right, or what the weather is like, or when I should expect a decision from *Whether*-q *Quarterly*. What I learn from the coconut isn't *relevant* to these issues, so why relitigate them? I have no new evidence, so I have little hope of coming to a better view. But I am certain to burn through precious time. No big deal for an unlimited thinker. But for efficiency-craving agents like me, it's better to be *selective* about when to update, and one good way to do that is to update only given changes in beliefs about things that are relevant.[16]

That's where disqualifiers come in. Since they make things irrelevant, we shouldn't

[16] For more on the costs of cognitive tinkering, see Holton 2014.

update given changes in already disqualified evidence. Caring about disqualifiers is thus part of being a selective updater; and the perk of selectivity is that our beliefs' bases tend to be more manageable. In this sense, disqualifiers bring 'epistemic upgrades'.

What kind of upgrades? Think back to Alice and Bob. When I only have Alice to go on, I have only one way to base my belief in *p*: I must base both on Alice and on Bob. Why? Because my *p*-belief has to be directly based in <Bob said *p*>, which I believe wholly based on <Alice said that Bob said *p*>. But once I see Bob's letter, I don't need Alice to know what he said, so I have the option of basing just on him—a better basing structure.

Sure, but what *makes* it better? One thing: basing on both would be intrinsically unfit. Alice's testimony is disqualified, and so no longer relevant to *p*. And whatever the 'basing relation' is, it ought to be true that one shouldn't base a belief on matters known to be irrelevant. After all, we report bases with explanatory locutions—'I believe that *because* of this', 'Here's the *reason* for which I believe it'—and irrelevancies can't explain.[17]

For limited reasoners, however, a more vital good-maker is that a basis with just Bob takes less time to manage: the trimmer my basis, the less potential for undercutters to make me rethink things. An illustration: Cate credibly warns me, 'Alice can't be trusted!'. Now Alice is

[17] But what *is* basing? I understand bases as correct, canonical answers to questions of the form, 'Why do you believe φ?', heard as a request not for straightforward causal or normative explanations, but for something more like a motive. But this is just a gloss. There are also four theories of basing: *causal* theories say that a basis for *p* is a proposition belief in which normally sustains belief in *p* (Moser 1989, p. 157); *counterfactual* theories say that the basis is what either did cause or would have (in certain cases) caused belief in *p* (Swain 1979; see also Korcz 2015, §2); *doxastic* theories say that basing involves or reduces to having meta-beliefs about what justifies what (Tolliver 1982; Longino 1987); and *causal-doxastic* theories say that basing requires either causal links or meta-beliefs (Korcz 2000). But theories aside, it's plausible that there is a link between what a belief is based on and when one must rethink it; for a nearby claim, see Lipson and Savitt 1993, p. 59 and Kelly 2002, p. 175 on the link between basing and revising.

undercut, and if her testimony is part of my basis for believing *p*, I am going to have to rethink my *p*-belief, because—even though I ought to keep it—I need to lop a proposition from its basis: <Alice said that Bob said *p*>. What a waste! Had I just based on Bob, hearing Cate wouldn't have triggered tedious revisions.

Still, won't there be *some* scenarios where I have to revise more if I base on Bob alone? It's tempting to think so. Suppose I hear Carl (also credibly) say 'Bob can't be trusted!'; now, if I've based only on Bob, I have to reconceive my relations to *p*, dropping my *p*-belief. But things would be just as bad if I had based on Alice, too. If Bob is undercut, so is Alice, because the path from her to *p* includes the path from Bob to *p*. Don't be tempted: a svelte, fully qualified basis really does mean less disturbance from defeaters.

So that's part one of the rationale: treat disqualified evidence as irrelevant, and on the whole, updates take less time. Part two has to do with space. In general, we ought to keep track of whatever is relevant to the propositions that interest us. (Forgetting evidence for trivia is fine.) But since our memory is finite, we shouldn't hoard every proposition that we stumble on. Again, it pays to be selective; and again, disqualifiers help slim down selections: if we just care about whether *p*, we may forget about Alice, and can make do with Bob.

Admittedly, forgetting Alice might be costly. Her testimony will be useful in cases that feature not just undercutters, but also their opposite—*uplifters*, which generate new evidential links. Suppose I've already heard Carl undercut Bob, when I hear Donna say 'If Alice says that Bob said *p*, then *p* is really true, regardless of whether you can trust Bob'. (Maybe she's privy to some more oblique Alice-*p* link.) Now Alice's testimony—at first irrelevant thanks to Carl—is 'uplifted' thanks to Donna, enough so to justify believing *p*. But I will only enjoy this boost in justification if I remember what Alice said. If I forgot her right after hearing from Bob, I am out of luck. Doesn't this show that there is some value in hanging on to evidence despite disqualifiers?

Yes, but there is *some* value in hanging on to any information, because anything *might* be uplifted into being evidence for nearly anything else. Instead of Donna, I could have heard from Deepak that *p* is true if there are 98 tiles on my kitchen floor. But even if I'd counted them last year, the act of committing the number to memory (just in case, against all odds, it wound up linked to *p*) would have been foolish. Just because something *might* become relevant doesn't mean it's actually relevant—or worth remembering.

So there is no real perk to a policy of basing in both qualified and disqualified evidence, whereas the costs are clear: wasted space and frittered time.[18] These are the first two parts of the rationale for taking disqualifiers seriously; both flow from a familiar meta-principle:

> *Principle of Clutter Avoidance* The principles of [belief] revision must be such that they discourage a person from cluttering up either long-term memory or short-term processing capacities with trivialities. (Harman 1986, p. 15)

A principle telling us to base beliefs even partly on disqualified evidence would clog our processing and memory; so, we should find another principle to live by.[19]

[18] One exceptional perk: having Alice's word is useful if I have accidentally *forgotten* seeing Bob's letter. (Forgetters love redundancy.) Clutter-reduction thus involves a tradeoff: one frees up space, but loses backup evidence. And the tradeoffs can get tricky, because 'being clutter' comes in degrees. (Say I'm 50/50 whether T3 tracks T1 or the weather, and I see both T3 and T1. T3 is *more* cluttering than T2 was, but *less* than T1; see §3, and recall the distinction between partial and total disqualification, from §1.)

[19] This might seem overblown. Is it really so hard just to keep track of Alice? Maybe not, but clutter is a bigger pain when there is more epistemic junk around. Recall my 26 sources on the state of *Zed's Head* (§1): Zed said he has a headache, Yuna said he said so, Xavier said Yuna said he said so, etc., all the way to Bob and Alice. Since I'm keen to know Zed's mental states, I suppose I could try keeping track of each

Aside from spatiotemporal clutter, there is a final reason not to base on disqualified evidence: *ease of use*. Suppose I'm wondering whether to believe H, and I directly know E and F, both of which entail it. But whereas the F–H link is obvious, the argument from E to H is mind-scrambling in length and complexity. Even if I somehow correctly followed it to its conclusion, the right response would not be comfortable belief in H, but 'rational self-doubt' (the term is from Christensen 2008; its application to hard arguments is in Schechter 2013). Since I know I struggle even with easier arguments, I should suspect that *somewhere* on the path from E to H, I messed up. So even though I should believe H, and even though I know E (which entails it), I should treat E as irrelevant; I ought to base just in F.

One way to explain this judgement is to say that F is an overwhelming disqualifier. F is manifestly decisive evidence for H, so it makes lesser evidence like E irrelevant—though notice that 'lesser' in this case means 'less fit for rational basing', not just 'less confirmatory'. (Contrast this with the guru case in §1; my vision confirms <The wall is red> *less* than the guru does.) We can also imagine a variation of the case where ideal, self-doubting agents should avoid basing on *subsumed* evidence; just suppose that E* supports H by virtue of supporting F in some unthinkably byzantine way. An agent who dimly sees but can't rationally trust the E–F link—and who knows that there is no other path from E* to H—would do well to treat E* as subsumed, and thus irrelevant to H.

So that is the three-part rationale for disqualification. If we take disqualified evidence to be irrelevant, refusing to base beliefs on it, we save time on revisions, we save space for what

utterance, making flash cards to remind myself, rethinking things whenever a link is undercut, resolving to do the same in a barrage of further cases—but then I would be the one with the headache.

matters, and we rely less on unwieldy evidence. The first two aspects reflect limits on our processing and storage; the third reflects our fallibility.

But, finally, even though this rationale comes from our non-ideal limits, there is a way in which disqualifiers could matter even to ideal agents. As David Christensen (2008) argues, even 'ideally rational agents'—whose beliefs are coherent, and whose reasoning is perfect—should often *believe* that they make mistakes. They aren't omniscient deities, after all: even perfect reasoners don't always know all the facts; and so they might not know their own infallibility, if there is enough misleading evidence around, and if their track record isn't yet absolutely impressive. To the extent that even these agents must worry about inferential gaffes, even they have to prefer shorter, safer paths from evidence to hypothesis—and therefore prefer fully qualified bases (like F) to ones containing obscurely related junk (like E) and subsumed distal indicators (like E*). If this is right, then disqualifiers matter for all but the most marvelous of reasoners: luminously perfect gods with bottomless storage and infinite time on their hands. Not much room for improvement, there.[20] But the rest of us need every shortcut we can get. [21]

[20] But note that, on some views, it would be an improvement if these agents didn't think of themselves as infallible! For Christensen (2008, §4) doesn't just argue that *some* ideal reasoners should *sometimes* doubt their abilities. He argues that it is hard to imagine how they could *ever* gather enough evidence to justify absolute self-confidence. So don't take my marveling too seriously.

[21] My thanks to audiences at MIT's MATTI reading group; MIT's undergraduate philosophy club; the 2017 meeting of the American Association of Mexican Philosophers; the ANU's Epistemology Reading Group; and the 2018 Central APA. I also thank David Balcarras, Nathaniel Baron-Schmitt, Selim Berker, Elise Bickford, David Boylan, Cherie Braden, Alex Byrne, Jonathan Dancy, Kevin Dorst, Jonathan Drake, Samia Hesni, Richard Holton, Meica Magnani, Clinton Neptune, Agustín Rayo, Becca Rothfeld, Raul Saucedo, Haley Schilling, Mike Titelbaum, Quinn White, Tim Williamson, and Michelle Xu. For detailed comments, I am grateful to Yael Loewenstein and Jack Spencer (who read this paper's unpublished ancestor), two anonymous referees, Joel Ballivan, Michael Bergmann, Moisés Macías Bustos, Nilanjan Das, Miriam Schoenfield, Ginger Schultheis, Kieran Setiya, Brad Skow, and David Builes. It was David who got me wondering about thermometers, and whose (relentless) objections pressed me for a rationale; I owe him special gratitude. Finally, I want to thank Steve Yablo for his kindness, patience, and endlessly inspiring papers on causation, all of which have made this project such a joy to puzzle through.

References

Atkinson, David & Peijnenburg, Jeanne 2013, 'Transitivity and Partial Screening Off', in *Theoria*, 79/4: 294–308.

Barnett, Zach forthcoming, 'Belief Dependence: How Do the Numbers Count?', in *Philosophical Studies*.

Baron-Schmitt, Nathaniel ms. 'Contingent Grounding.'

Christensen, David 2008, 'Does Murphy's Law Apply in Epistemology? Self-Doubt and Rational Ideals', in *Oxford Studies in Epistemology* 2: 3–31.

Dancy, Jonathan 2004, *Ethics Without Principles.* Oxford: Oxford University Press.

Elga, Adam 2007, 'Reflection and Disagreement', in *Nous*, 41/3: 478–502.

Fitelson, Branden 2012, 'Evidence of Evidence is Not (Necessarily) Evidence', in *Analysis*, 72/1: 85–88.

Goldman, Alvin 2001, 'Experts: Which Ones Should You Trust?', in *Philosophy and Phenomenological Research*, 68 (1), pp. 85–110.

Harman, Gilbert 1973, *Thought*. Princeton: Princeton University Press.

————1986, *Change in View*. Cambridge: MIT Press.

Holton, Richard 2014, 'Intention as a Model for Belief', in *Rational and Social Agency: The Philosophy of Michael Bratman*, Eds. Vargas, Manuel and Yaffe, Gideon. Oxford: Oxford University Press. 12–37.

Johnson King, Zoë forthcoming, 'We Can Have Our Buck and Pass It, Too', in *Oxford Studies in Metaethics*, vol. 14.

Kagan, Shelly 1988, 'The Additive Fallacy', in *Ethics*, 99/1: 5–31.

Kamm, Frances Myrna 1983, 'Killing and Letting Die: Methodological and Substantive Issues', in *Pacific Philosophical Quarterly* 64/4: 297–312.

Kelly, Thomas 2002, 'The Rationality of Belief and Some Other Propositional Attitudes', in *Philosophical Studies* 110: 163–96.

————2005, 'The Epistemic Significance of Disagreement', in *Oxford Studies in Epistemology, Volume 1*, eds. Hawthorne, John and Gendler, Tamar, pp. 167–196.

————2016, 'Evidence', in *Stanford Encyclopedia of Philosophy* <https://plato.stanford.edu/archives/win2016/entries/evidence/>.

Korcz, Keith Allen 2000, 'The Causal-Doxastic Theory of the Basing Relation', in *Canadian Journal of Philosophy*, 30/4: 525–550.

————2015, 'The Epistemic Basing Relation', in *Stanford Encyclopedia of Philosophy* <https://plato.stanford.edu/archives/fall2015/entries/basing-epistemic/>.

Kotzen, Matthew ms., 'A Formal Account of Epistemic Defeat', unpublished manuscript.

Kripke, Saul 2011, 'On Two Paradoxes of Knowledge', in his *Philosophical Troubles: Collected Papers, Volume I*. Oxford: Oxford University Press. 27–51.

Lackey, Jennifer 2014, 'Disagreement and Belief Dependence: Why Numbers Matter', in *The Epistemology of Disagreement: New Essays*, eds. Christensen, David and Lackey, Jennifer. Oxford: Oxford University Press, pp. 243–268.

Lasonen-Aarnio, Maria 2014a, 'The Dogmatism Puzzle', in *Australasian Journal of Philosophy*, 92/3: 417–32.

————2014b, ‘Higher-Order Evidence and the Limits of Defeat’, in *Philosophy and Phenomenological Review*, 88/2: 314–45.

Lehrer, Keith & Paxson, Thomas 1969, ‘Knowledge: Undefeated Justified True Belief’, in *Journal of Philosophy*, 66: 225–37.

Lipson, Morris and Savitt, Steven 1993, ‘A Dilemma for Causal Reliabilist Theories of Knowledge, in *Canadian Journal of Philosophy*, 23/1: 55–74.

Longino, Helen 1978, ‘Inferring’, in *Philosophy Research Archives*, 4: 19–26.

Moser, Paul 1989, *Knowledge and Evidence*. Cambridge: Cambridge University Press.

Muñoz, Daniel ms. ‘Echo Chambers and Dependent Evidence’.

Pollock, John 1970, ‘The Structure of Epistemic Justification’, *American Philosophical Quarterly*, monograph series 4: 62–78.

————1974, *Knowledge and Justification*. Princeton, NJ: Princeton University Press.

Pollock, John & Cruz, Joseph 1999, *Contemporary Theories of Knowledge*, 2nd Edition. Lanham, MD: Rowman and Littlefield.

Reichenbach, Hans 1956, *The Direction of Time.* Dover Publications.

Schechter, Joshua 2013, ‘Rational Self-Doubt and the Failure of Closure’, in *Philosophical Studies*, 163/2: 428–452.

Shogenji, Tomoji 2003, ‘A Condition for Transitivity in Probabilistic Support’, in *British Journal for the Philosophy of Science* 54/4: 613–616.

Skow, Bradford 2016, *Reasons Why*. Oxford: Oxford University Press.

Sudduth, Michael, ‘Defeaters in Epistemology’, in *Internet Encyclopedia of Philosophy* <http://www.iep.utm.edu/ep-defea/>

Swain, M. 1979. ‘Justification and the Basis of Belief’, in *Justification and Knowledge*, Ed. Pappas, George. Boston: D. Reidel.

Titelbaum, Michael ms., *Fundamentals of Bayesian Epistemology*.

Tolliver, Joseph 1982, ‘Basing Beliefs on Reasons’, in *Grazer Philosophische Studien*, 15: 149–161.

Weatherson, Brian ms., Do Judgments Screen Evidence?

Yablo, Stephen 2010, ‘Advertisement for a Sketch of an Outline of a Prototheory of Causation’, in his *Things*. Oxford: Oxford University Press. 98–116.

Zagzebski, Linda Trinkaus 2012, *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*. New York: Oxford University Press.