Abstract: Justification and the Growth of Error

Most epistemologists, externalists and internalists alike, have accepted that the justification of belief is fallible, but appreciating and incorporating the implications of this is an ongoing process. One of these implications is that potential error grows with every fallible step. I argue that this worrying conclusion is not directly in conflict with common sense. However, it does appear to undermine the idea that we can improve our epistemic situation by taking extra steps of several familiar sorts, such as reliability-checking, proof-checking, and gathering more evidence. I explain how error grows over steps, and thereby why this skeptical conclusion is not forced on us.

Justification and the Growth of Error[1]

Most epistemologists, externalists and internalists alike, have accepted that the justification of belief is fallible, but appreciating and incorporating the implications of this is an ongoing process. One of these implications is that potential error grows with every fallible step. I will argue that this worrying conclusion is not directly in conflict with common sense. However, it does appear to undermine the idea that we can improve our epistemic situation by taking extra steps of several familiar sorts, such as reliability-checking, proof-checking, and gathering more evidence. I will explain how error grows over steps, and thereby why this skeptical conclusion is not forced on us.

A Starting Point: the Connection Thesis

The Connection Thesis is intuitively compelling: If you are justified in believing there is a reliable information source according to which p, then you are justified in believing p. If I am justified in believing that the car mechanic is a reliable source, and the mechanic says I need a new battery, then I am justified in believing that I need a new battery. The interest of the thesis lies in the fact that it looks as if it is not true on any externalist view of justification. (Vogel 2006) I will argue that the Connection Thesis is not true on any view of justification that takes justification to be fallible, whether externalist or internalist. Thus, the thesis does not provide a wedge between the two paradigms. It does highlight questions that fallibility raises for any view of justification due to the fact that error grows with every fallible inference. In particular, the growth of total potential error that comes with every

1

additional premise or step of inference appears to imply, paradoxically, that acquiring more information about the reliability of our sources and checking our proofs makes us less rather than more justified. It also appears to imply that taking into account more evidence makes our empirical conclusions less justified. In this paper I identify how error grows over additional premises and inferences, including those that involve claims *about* reliability, and offer a resolution of this paradox.

The Connection Thesis can be represented:

$J(R(p)) \rightarrow J(p)$

If I am justified in believing there is a reliable information source according to which p, then I am justified in believing p. Take the representative externalist theory to be pure reliabilism, that is, reliabilism without any internal requirement. Then, roughly, a belief is justified if and only if it was due to a reliable information source:

$J(p) \leftrightarrow R(p)$

Thus for reliabilism to imply the Connection Thesis would be for it to imply the following:

$R(R(p)) \rightarrow R(p)$

If I have a reliably formed belief that I have a reliably formed belief that p, then I have a reliably formed belief that p. But this claim is false. Rewrite it, representing R(p) as q:

$R(q) \rightarrow q$

Clearly there is no extant externalist theory on which having a reliable belief that q implies that q is true. The belief's being reliable implies at most that q is probably true. It is tempting to think of this as a consequence of externalism itself. That one has a reliable (or safe, or tracking) belief that p is a separate matter from whether p is true. The reliability of one's belief that p and p's truth are two distinct facts about the world, and the one cannot simply collapse to the other. By contrast, reliability claims appear to have a mysterious transparency for the intuitive notion of justification that is highlighted by the Connection Thesis. It does not matter whether the reliability claim about a source is actually true; if I am justified in believing it then that justifies me in believing what the source says.

However, the failure of the Connection Thesis for externalist theories may just be a consequence of fallibilism, the fact that the reliabilist view takes it as possible to be justified while one still might be wrong. Fallibilism is definitely assumed in the argument that the Connection Thesis is false for externalism, as is natural because we never meet infallible belief-forming processes so externalists never raise the bar for justification that high. It is because the reliability expressed in "R(q)" is fallible that R(q) does not imply q, the crucial step in the argument above. We can confirm this by considering an infallibilist, externalist criterion for justification, which is of course logically possible. On this view, justification requires infallible reliability. That is:

$R(q) \rightarrow q$

If a belief is infallibly reliable, then the belief is true. Substituting for q the expression "R(p)," we get the Connection Thesis for externalism:

$R(R(p)) \rightarrow R(p)$,

which says that if I have a perfectly reliably belief that there is a perfectly reliable source according to which p, then I am perfectly reliable in believing p. For an infallibilist reliabilist, the Connection Thesis is true; perfect reliability is preserved under composition. Externalist justification is transparent too, if we crank it up to an infallibilist level. Since the Connection Thesis is false for externalism when reliability is assumed fallible, and true for externalism when reliability is assumed infallible, the interest of this thesis has at least as much to do with fallibility as it does with internalism and externalism.

Attrition for all

We can make the Connection Thesis true for reliabilism if we imagine reliabilism as an infallibilist view. But intuitive justification is fallible just like standard reliabilist justification is. Can we also make the Connection Thesis intuitively false by focusing on the fallibility of intuitive justification, now going back to the more natural assumption, for R(x), that a belief-forming process can be reliable without being infallible? The answer is affirmative, and the reasons are illuminating about fallibilist justification in general.

Consider again the mechanic who told me that I needed a new battery. That I am justified in believing that he is a reliable source about my battery, and that he is a reliable source, are distinct facts. If the first is true, then I have good reason to believe that what he reports about batteries is usually correct. If the second is true then what he reports about batteries is indeed usually correct. But, obviously, on the assumption that intuitive justification is fallible the first does not imply the second. That is because my good reason to believe only takes me to a high level of justification, not to a level that implies the truth of the believed proposition. Let us say that level of justification is 95%. The level of reliability I attribute to the mechanic is of course a different matter. We can imagine that what I have reason to believe *about* him, represented by “R(p)”, is that he is 90% reliable. That is, I have a 95% level of justification for believing that 90% of his reports about batteries are true.

It is a subtle matter what level of confidence about my need for a battery I have a right to in this case. (More about this below.) But whatever the calculus should be that combines my level of justification for believing he is reliable with the level of reliability I attribute to him, it seems clear that the resulting level of my justification for believing I need a new battery is not 95%. What I have a 95% right to believe is not that he is right, but that he tends to be right 9 out of 10 times. This cannot give me a 95% right to believe p. That is, the property indicated by the first "J" in the Connection Thesis

$$J(R(p)) \rightarrow J(p)$$

is not the same as the property indicated by the second "J". The property the second "J' refers to does not necessarily fail to be any kind of justification. Indeed, in this case, it surely will still be a pretty strong one. But the fact that two steps were made to get to belief in p, and that both steps gave me reason to believe my conclusion might be wrong, means that the resulting level of justification is lower, however minutely, than the level of justification of the first judgment, so the two "J"s refer to different properties. The Connection Thesis is not generally true, even intuitively. It seems true partly due to underdescription. This is so even on an internalist conception of justification according to which justification is a matter of having reasons, since I put no restriction on what "justified" means here except that it comes in levels, that is, that some justifications are stronger than others.

It is also clear that the assumption that yielded this type of counterexample was fallibilism. If we assume that intuitive justification and reliability are infallible, then the Connection Thesis is true. For consider that in that case my being justified in my belief about the mechanic's reliability implies that the mechanic indeed has the reliability I attribute. And since that reliability is also perfect by assumption, then what he reports is indeed true. My infallibly justified belief that the mechanic is infallible, gives me an infallibly justified belief that I need a new battery. That is, there isn't a node in this cascade that introduces any potential for error, so my justification for believing I need a new battery is infallible too. So, the first "J" in the Connection Thesis indicates the same property that the second "J" does.

But this argument works just the same if "J" refers to a fallible sort of justification. The assumed infallibility of the "R" property means that the level of justification indicated by the first "J" remains in the second "J," whatever that level may be since the R claim introduces no possibility of error. However, the fact that the Connection Thesis can be made true in this latter way, that allows fallibility of the intuitive justification, doesn't help the internalist against the externalist, since, ironically, the Connection Thesis only turns out true here because the externalist's criterion, reliability, is assumed *in*fallible.

We have seen that if both J and R are fallible, then the Connection Thesis is false, and that if J and R are both infallible, then the thesis is true. We have seen that if J is fallible and R infallible, then the Connection Thesis is true but not useful. It remains to consider the last mixed fallibilist/infallibilist case. This is where R(p) does not imply p, but J(p) does imply p. Suppose the level of reliability in R is 90%. If we accepted the reasoning above, then it seems clear that the Connection Thesis is false:

J(R(p)) does not imply J(p).

The first J here is assumed infallible. Thus,

J(R(p)) implies R(p)

But of course R(p) does not imply p, since R is fallible. Still, the intuition goes, we are *justified* in believing p! However, even if we have justification, the Connection Thesis is not true in this case, since for it to be true the predicate J referred to by the first “J( )” must be the same as the J referred to by the second occurrence of “J( )” of the thesis. That first J is infallible, and if we assert p on the basis of J(R(p)) then even though we seem to be justified we cannot be *infallibly* justified in asserting p, due to the fallibility of R, so the sense of justification is clearly different. The level of confidence in p that you have a right to in this case is strictly less than the level of confidence in R(p) that you had a right to. Even if the justification of the belief that there is a (fallibly) reliable source is infallible, the Connection Thesis is false.

Cases where the Connection Thesis (CT) is false are all and only the ones where we have taken the reliability claim R( ) to be fallible. Accepting this assumption, the Connection Thesis isn’t true even if justification, J( ), is internalist and infallible. Vogel is well aware that what makes the Connection Thesis false for the externalist is that the reliability claim is fallible. (Vogel 2006) The news is that this makes the thesis false for internalism too, no matter what kind of internalist you are. The internalist might have hoped that the falsity of CT for externalism and its truth for internalism would show that whereas reliabilist justification exhibits attrition over added steps in reasoning, internalist justification does not. A threshold conception of justification which depends on how reliable a belief formation process is, is widely assumed to be the source of this problem: enough fallible steps will put one’s total potential error below the threshold of error-avoidance required for justification. Such a conception seems to be forced upon the reliabilist, who in defining justification in terms of reliability seems to be forced to define reliability by frequency of true beliefs, and infrequency of errors, and so by levels. A threshold conception is not forced on the internalist, it is presumed, so she need not deal with this problem.

However, the source of the attrition of justification over steps is not a threshold conception of justification, where a belief is counted as justified or not depending on whether its credentials rise to a particular level. The source is rather the conception of justification as coming in levels at all. And the internalist must also grant that justification comes in levels if she grants that some justifications (of a given p) are stronger or weaker than others. The issue is about strengths of justification, not thresholds, because what mattered in the arguments above was not that the change from the first “J( )” to the second “J( )” was a change that brought the justification below a particular magic number, but that it brought the justification down at all, that it changed, weakened, the justification property. The strength of internalist justification can be expected to drop off in just the same way as we have seen for reliabilist justification.

If we acknowledge the fallibility of justification, be it of the internalist or the externalist variety, we must admit its deterioration over the inferences discussed, where a justified reliability judgment about a source is followed by an inference to p. But the point appears to be general over all sequences of inference with multiple steps. If what is concluded above is true, then it does not seem deniable, if every step in a proof has non-zero fallibility, that every additional step will increase the potential error and thereby weaken the justification of the concluding belief. However, these two types of sequences of inference do have different structures—the first is a composition of a justification claim with a reliability claim via a “that” clause, while the second is a concatenation (or conjunction) of fallibly justified inferences—and in what follows we will be more careful about the differences.

If the Connection Thesis is wrong, it is fair to wonder why it looked so right. This, I think, is for at least three reasons. One, as above, is that we had underdescribed both the situation and what intuitive justification is like. Reliabilism or externalism, as a challenger to received wisdom about justification, has had to be defined explicitly in order to be a contender, so the failure of the Connection Thesis for it is more easily derived. Second, the phenomenon that makes the CT false is the attrition of justification over steps, and, as we will see below, this attrition is increases with the number of steps. Since the CT involves only two steps, the attrition is small. It is harder to notice intuitively than it would be if several compositions were involved. A third reason is that a weaker version of this thesis *is* true. The weaker version says that though in composing a justified belief in the reliability of a source with what the source reports there will typically be some weakening of justification, still there is a lower bound on how much the loss will be, a bound related to the two levels of justification going in. Whatever the calculus might be, there is no way that combining a 95% level with a 90% level will bring you to no justification at all, or to a random level of potential error.

Because the difference between the false versions and the true version of the Connection Thesis is quasi-quantitative, and does not make a difference in many or most common cases, ordinary usage of the word "justified" does not force one to observe it. In the case considered above, though my level of justification for believing in the reliability of the mechanic is 95% and the resulting level of justification for believing *what* he has said must be lower, it is not very much lower. If it is still 85% or above, say, there is no point in our noting the difference in our reports of the situation in ordinary life or language, since one is still pretty well justified to believe what he says. If having a justified belief means having a belief one has a right to, then, broadly speaking, a belief's being "justified at level x" and "justified at level x minus 5%" will both intuitively count as justified or unjustified depending on whether the two levels fall on the high side or on the low side of the spectrum. If they fall in the borderline area of the spectrum of well-justified and not justified, then it's unclear whether they count as justified or unjustified, but here too the feeling we tend to will probably be the same for both. This is why we don't observe the quantitative distinction, and that is one of the reasons we are misled into thinking the Connection Thesis is true. I will discuss below further reasons why we should not be surprised that intuitive counterexamples to the Connection Thesis are hard to construct.

In the next section I will make a proposal about how the levels of fallibility of justification and of reliability claims should be combined to yield a level of reliability for the resulting belief. This will provide an understanding of how quickly or slowly error grows in general for chains of inferences. But because it is so clear that, for example, a 95% level of justification combined with a 90% level of claimed reliability will not leave you *un*justified, such a fact can also be taken as a condition of adequacy for that proposal.

How much, how fast?

Imagine an example of the inference referred to by the Connection Thesis. The reliable source could be a thermometer, but here we will let the source be a person, for ease of exposition. Suppose you form a belief that p on the basis of the fact that Mrs. Jones told you that p. If we asked you why you believed p, this is what you would refer to, and if we accepted that then we would be taking you to be justified through one step, for one reason. Now imagine a case where you arrived at your belief that p slightly differently. After hearing what Mrs. Jones said, you checked with many people who know Mrs. Jones to see if she is reliable when it comes to matters like p. Now you believe p and your reasons are that Mrs. Jones told you so, and Mrs. Jones is reliable. How does the added reliability belief change the level of justification of your belief that p?

It seems that there must be a contribution from both the level of justification of the reliability belief, and the level of reliability of Mrs. Jones that that belief reports. An obvious candidate is the idea that the added reliability belief enhances justification when the content of the belief—e.g. "Mrs. Jones is 70% reliable"—if true, reduces the potential error of the overall inference more than the step of believing that reliability claim adds to the overall potential error. Attrition of justification would occur in the reverse case, where the level of (un)reliability with which you believe the, say, 70% reliability claim about Jones introduces more potential error than the claim, if true, would eliminate.

But this cannot be right. If in checking on her you instead asked random passersby on the street whether Mrs. Jones is reliable and you believed she is reliable because most of them happened to say yes, your potential for error in believing on this basis that she is reliable would be very high. You are close to 50% in the probability of being right about her, and we can even suppose that this is obvious to you.[i] Since this is the maximum potential error, it surely does not matter what level of reliability the testimony told you to attribute to her. To put the point in terms of internalist-style justification, even if they told you she was perfectly reliable, the fact that you chose these people randomly means their testimony gives you no reason at all to believe that. In adding the reliability belief on this basis, and adding this reliability belief as a basis for you final conclusion p, you appear to have maximized the overall potential error of your resulting belief in p, regardless of what your level of justification for your initial belief in p based on Jones's testimony without a reliability belief was.

This obvious maxing out gives a clue to how Mrs. Jones's putative reliability in judging p and the reliability of my judgment of her reliability should be aggregated to give the reliability, or justification level, of my resulting judgment p. Consider the following statements, changing "you" to "George" to avoid an implication that the person believing the source knows anything about his or her own state:

$R_s(p) = x$ if and only if Mrs. Jones is x% reliable in judging p.

$R_1(R_s(p)) = y$ if and only if George is y% reliable in judging the level of reliability of Mrs. Jones (about p).

The endpoints, where each of these involves either perfect randomness[ii] or perfect reliability of your judgments, are clear:

"$R_s(p)$"	$R_1(R_s(p))$	$R_2(p)$
x	y	final level
1	1	1
0	0	0
1	0	0
0	1	0

In the first case, George judges that Mrs. Jones is perfectly reliable in her testimony that p, and he is perfectly reliable in judging that, so he ends up with perfect reliability in judging p.[iii] In the second case he is at random when he judges that his source, Mrs. Jones, has zero reliability (is at random) in judging p. Therefore, he has zero reliability if he believes p on the basis of her testimony. Similarly if George is at random when he judges that Mrs. Jones is perfectly reliable on p, then this reliability belief does him no good. If he believes p on the basis of this encounter, he has no reliability in doing so. Finally, if George is perfectly reliable in judging that she has zero reliability (is at random) in judging p, then his encounter gives him zero reliability in judging p.

The Connection Thesis studied in the chart above is the version for reliabilism, $R(R(p)) \rightarrow R(p)$. We must verify that the same holds for the attrition of intuitive justification as holds for reliabilist justification.

"$R_s(p)$"	$J_1(R_s(p))$	$J_2(p)$
x	y	final level
1	1	1
0	0	0
1	0	0
0	1	0

In the first case, George judges that Mrs. Jones is perfectly reliable in her testimony that p, and he is perfectly justified in judging that, so he has perfect justification in judging p. In the second case George has zero justification in judging that his source, Mrs. Jones, has zero reliability (is at random) in judging p. Therefore, he has zero justification if he believes p on the basis of her testimony, since he doesn't know anything about her.

Similarly if George has zero justification for judging that Mrs. Jones is perfectly reliable on p, then the belief in her reliability does him no good. If he believes p on the basis of this encounter, he has no justification to do so. Finally, if George is perfectly justified in judging that she has zero reliability (is at random) in judging p, then his encounter gives him zero justification for judging that p.[2]

The most obvious general relationship to extrapolate from these endpoints is that the resulting reliability (justification) level in judging p is the *product* of the source's judged reliability level and George's reliability (justification) level in judging the source's reliability. If either ingoing level goes to random (zero), then the reliability (justification) of his judgment p goes to random (zero). Zero times anything is zero. If both levels are perfect, then the reliability (justification) of his judgment p is perfect (1 multiplied by 1 equals 1). Reliability and justification levels are seen as fractions between zero and one inclusive. The product of two such fractions is less than or equal to each of them, so the proffered relationship implies that if reliability (or justification) for believing p is built from sources that have a certain level of reliability and judgments about those sources' reliability that themselves have a certain level of reliability (justification), then the reliability (justification) of the resulting belief will always be less than either of the levels of the component judgments, unless one or both of the levels of those component judgments is perfect and none is zero. In no case can the resulting reliability (or justification) be greater than the lower of the two levels of reliability or justification going in.

[2] The correspondence between intuitive justification and reliability here is not offered as an argument that reliability is the right analysis of justification. It is not surprising that when we think of both in terms of levels, with a maximum and a minimum, this correspondence should arise. That the two are pro-epistmic properties and this feature about levels are what is at work in these points.

In the kind of case imagined we are calculating the resulting reliability (justification) level for a two-step process involving composition of "that" clauses: George believes, with a certain level of reliability, that Jones has a certain level of reliability about p. But the table above about the endpoints in the calculation of resulting justification or reliability applies equally to a case where we imagine conjoining claims, none of which need be about reliability. For "Our dogs are white" inferred from "Fluffy is white" and "Sport is white" would be a belief with maximum potential error if either of the premises had maximum error, and zero potential error if both premises had zero potential error (provided the inference was infallible, an issue we will return to). So we have just as much reason to think that the rule for figuring the resulting level of reliability or justification of an inference from conjuncts to conjunction is to multiply the reliability or justification levels of the conjuncts. This rule for error growth in conjunction inferences has been known for some time.

This all appears to be very bad news, since not only is there attrition of justification, but since the attrition goes as the product it goes pretty fast: if x, your level of reliability in judging Mrs. Jones, is .75, and y, the level of reliability you attribute, is .75, then the reliability of your judgment that p is .56, almost random. However, though our glass is half empty it is also half full, since the product rule for figuring the accumulating error puts not only an upper bound on justification or reliability of the resulting judgment, but also a lower bound on the growth of potential error. If you are 90% justified in your judgment of Mrs. Jones's reliability, and what you judge is that she is 90% reliable, then these facts would not make you any less than 81% justified in your resulting judgment that p. If you are 95% justified in your judgment of Mrs. Jones's reliability, and what you judge is that she is 95% reliable, then these facts would not give you any more than 10% of potential error in your resulting judgment that p.

That reliability drops off in this way over fallible steps is quite in line with common sense, and something we live with. Anyone who has played the game "Telephone," where a message is whispered from one person to the next along a line of, say, twenty people, will not be surprised that reliability drops off: the message coming out the end of this line typically bears almost no relation to the message that went in. We don't, I think, tend to trust messages that come through many middlemen, unless we are highly confident of the reliability of the links. The journalist is expected to interview the scientist directly, for example, if he wants to write a story about her discovery. And even there we are mindful that there will be some degree of mismatch due to the journalist's lack of expertise in the field, and the need to make the material comprehensible to the layperson reader, even though it is not a mismatch we are likely to be able to locate for the same reason. In a court of law, too, testimony to the effect that p that is based on hearing someone else say that they witnessed p, also known as "hearsay," is not allowed. On the other hand, we tend to trust that a photograph of a painting looks just like the original, even if the photo passed through the hands of many people, such as a museum photographer, an editor, a layout designer, a printer, and a bookdealer. Photographs are faithful to a high degree unless doctored, and museums and publishing houses have very high incentives not to doctor—such things are easily checked.

With a rule in hand for determining resulting justification level, namely, by multiplying the judged reliability level of the source, x, by the level of justification of that judgment, y, we can now see why clear counterexamples to the Connection Thesis are hard to produce via intuitions. The fact is that if this is the proper way to determine the resulting level, then the higher the justification level (y) and reliability level (x) going in to the judgment p, the lower in absolute terms the drop in the final justification level of the belief in p. For example, assuming judged reliability has the same level as the justification of that judgment, (.9)(.9) = .81, a nine-point drop, but (.7)(.7) = .49, a 21-point drop. The lower you start in justification and content of the reliability claim, the farther you fall in one step. The higher you start, the less you fall. Mapping these trends onto the binary justified/unjustified choice, the more inclined we are to call the initial judgment of reliability justified (.9) the less likely it is that the resulting judgment of p will seem *un*justified (.81). On the other hand, the more inclined we are to call the resulting judgment p unjustified (.49), the lower the level of justification had to have been going in (.7). It will have fallen farther, but its initial level still can't have been high. So we can't expect to see intuitive examples that go from seeming clearly justified in the first step (y) to clearly unjustified in the result, or examples that are clearly unjustified in the resulting belief p but clearly justified in the first step.

Our best bet is to look on the lower side where the drop is largest. It should be possible to find examples where the judgment of reliability (with level y) seems fairly good (.7) and the reliability attributed seems fairly good (.7), but the resulting belief in p seems clearly not good enough to be called "justified." These can be constructed. Imagine your friend John has a pretty good memory. You know from experience that he's right about three quarters of the time. Suppose you need a car mechanic and John tells you that the guy down at Ace repair did right by him a lot of the time. He has gone to the guy twice a year for some years and only about three times did he screw things up. Suppose you form the belief, p, that you should go to the Ace guy, on the basis of what John told you about the guy's reliability. Your belief about the mechanic's reliability is as good as John's memory, that is, three quarters justified. The content of that belief is, say, that the mechanic is three quarters reliable. But when you put the two together, you have: John's fairly reliable, and he says that the mechanic is fairly reliable, and it does not look like you are really justified in believing the mechanic will fix your car right. Conformably, the reliability level of your resulting belief would be 56%.

If one hesitates about this conclusion this is, I think, because it can be hard to put these two judgments in question together intuitively, that John is 75% reliable, and his report that the mechanic is 75% reliable. This is because when we take John's word, we believe and go with what he says or we don't. The question of how *far* to trust him appears, practically at least, no longer to apply. If John says the mechanic is 75% reliable, and I'm forming my belief by taking his word for it then I seem to be 75% justified in the belief the mechanic will fix my car right, right? Well, no, because in saying this I seem to have forgotten that John was only 75% reliable in what he said, and if my knowledge of that was the basis for my belief that the mechanic is 75% reliable, then how justified could that belief have been? Not 75%, surely. To see that the imperfect reliability I attribute to John matters, compare the case where I think his memory is perfect. I would certainly be better off in my justification level for believing the mechanic will fix the car right, than I was in the case as first described. Thinking carefully about the example, and seeing that there are two fallible parts to forming the judgment that the mechanic will fix your care right, makes it hard to deny the conclusion that justification slipped away here.

More steps, more worries

The idea that distinct steps along the way to a conclusion introduce potential error that grows with the number of steps is familiar from as far back as Hume's discussion of demonstrative proofs. (Hume 1978, Jeshion 1998**).** Be the rules of reasoning however infallibly valid—if A and B are true then, necessarily, B is true, for example—our application of these rules is not. We may apply them to the best of our ability and still proceed from a truth to a falsehood. Hume took this to imply an obligation to provide proof checks for each step of a demonstration, and at times took this to imply a skeptical regress, since it seems that the checks would themselves have to be checked to do any good. More recent discussion of the supposed fallibility of mathematical knowledge has reached a fair amount of consensus that proof-checking is required for justification. (Detlefson et al. 1980, Kitcher 1984, Resnick 1989. But see Jeshion 1998 for dissension.) If even demonstrative knowledge is fallible, and the potential for error grows with the number of steps, then it should not come as a surprise that potential for error in empirical knowledge can grow with the number of steps. (More below about what counts as a step.) It also appears to be unnecessary to make special internalist or externalist assumptions to see the problem for demonstrative reasoning, as I have just argued for empirical knowledge.

Nevertheless there is an apparent paradox from the attrition of justification that what is similar in the cases of proof-checking and reliability-checking highlights nicely. Suppose, as above, that you have justification for believing p and that justification goes through one step. You have a reason to believe p and it is that Mrs. Jones told you that p. If we asked you why you believed p, you would say that Mrs. Jones told you, and if we accepted that then we would be taking you to be justified through one step, for one reason. Now imagine a case where you arrived at your belief that p slightly differently. After hearing what Mrs. Jones said, you checked with many people who know Mrs. Jones to see if she is reliable when it comes to matters like p, and they told you she was. Now you believe p and your reasons are that Mrs. Jones told you so, and Mrs. Jones is reliable. Your reasons are two, or more, and they are interlocking, like steps of support. The apparent paradox is this: if potential error grows with the number of steps in the justificatory sequence, as it must if the calculation goes by a product and all reliability levels are between zero and one exclusive, then you appear to be *less* justified in the case where you appear to have *more* information and justification.

There is a similar seeming paradox with mathematical proofs: the step of checking a step in a proof is itself fallible, so though it seems that this checking is giving you more information, it is also introducing error. The product rule implies that in adding a source of error—such as an additional fallible premise—and doing nothing else, you strictly increase your potential error. So it seems that you will have strictly more total potential error when you have added a safeguard than when you have not, undermining the purpose of the safeguard. How can an added check on our source's reliability or on our inference make us worse off with respect to error? Or, given that we now know the way that potential error accumulates, how can it be, as we think, that adding a check makes us better off with respect to error? Note that this is different from the skeptical problem Hume raised about proof-checking. The problem he noticed was that in order for a check to help bolster the original inference it seems the checking statement, say, "that's a valid inference," would itself need to be justified, and if a check was required to justify the original inference why wouldn't the same be required to justify this one? There would have to be a check on it, and a regress follows. Here the problem is not how to justify the check, but that however we assume that the checking statement, "that's a valid inference," is justified, if the justification is fallible then it seems it cannot do the job of taking away potential error, but only add to the potential for error.

This can of worms is even bigger, for notice that as we gather more empirical evidence for a conclusion, we add more and more premises, each of which is fallible. Our conclusion should thus become less reliable and less justified, because of the growth of error with an increasing number of premises. How, then, is it possible for us to make our conclusions more secure by taking more evidence on board?

One obvious point to make about growing error is that the number of steps is not the only factor in how much potential error there is in your resulting belief. Different kinds of steps have different levels of potential error. Some judgments of reliability (as of Mrs. Jones) are harder to make or support than others are. Some steps of reasoning are more difficult than others. In addition, some steps of reasoning cover more ground than others. It is conceivable that going through each and every step of a lengthy "gap-free" proof has less potential for error than reasoning in one step from the premise to the conclusion. Indeed, to some this has seemed obvious, and was one of the primary motivations for developing formal systems. The reason the gap-free proof is less error-prone, if it is, has to be that the small steps are so clear and easy to see, and thus the potential for error in each one of the small steps so tiny, that even though there are many steps the resulting total of potential error is also tiny. By contrast, the one big step from the starting premise to the same conclusion can be a jump over a matter that is complex, difficult and obscure, with these features contributing to a large potential for error. We can confirm this comparison with the product formula for accumulating error: five steps with 99% reliability each will get you significantly more safely to your conclusion than will one step with 90% reliability, since $(.99)^5$ is .94. If there is a paradox from the attrition of justification, it does not appear to be general. For one thing, justification is not always weaker when it results from a greater number of steps.

A puzzle remains even with this kind of case, though. When we fill in the gaps in demonstrative proof, don't we *add* steps as we interpolate between A and D (where A implies D) inferences from A to B, B to C, and C to D (where each implies the next)? If so, then it doesn't matter how small the error in the added steps is. Since they are in addition to the steps we started with, it seems they have to bring down the total reliability; so says the product formula. This particular puzzle is easily solved. Filling in the gaps of a proof is not a matter of adding steps, although it certainly seems so, perhaps because the proof we end up with has strictly more steps. However, the inference in the initial proof is not among the inferences in the revised version of the proof, so what has occurred is not an addition. We begin with an inference from, say, A to D, and we end with inferences from A to B, B to C, and C to D. The inference from A to D does not exist in the gap-free version of the proof. Filling in gaps of a proof is capable of enhancing our justification for asserting the conclusion, in part because this procedure is not a matter of addition of steps, but of replacement of the original proof with a new proof.

The case of checking up on Mrs. Jones or checking a step in a proof is not settled by this point, though. Here it is not that we augment steps by interpolating between A and D (where A implies D) inferences from A to B, B to C, and C to D (where each implies the next), as we do when filling in the gaps of a demonstrative proof. That Mrs. Jones said p does not by itself imply or even make more probable that Mrs. Jones is reliable, so in adding a belief about her reliability we are not drawing out an intermediate conclusion from our first premise. Rather, the reliability claim supports the inference from the fact that Mrs. Jones said p to p. Similarly, checking a step of proof is not making explicit an intermediate stage in the inference from one line to the next, but rather supports the inference from one line of the proof to the next through a judgment that the inference is valid. In both of these cases we are adding a premise to the proceedings which carries its own level of potential error, however small. The product formula says adding steps or premises makes our justification level worse, not better.

Seeing this process not as an addition to an existing inference but as creating a replacement for the original justification will not work here. Appealing to the premise about Jones's reliability was indeed the addition of a conjunct; we formerly had only the premise "Jones said p" and now we have that plus "Jones is 90% reliable." The error grows with the number of fallible conjuncts, and regarding this as a new inference won't change the fact that merely adding a premise cannot make it a safer inference, and usually makes it a more perilous one. Yet surely the addition of a highly justified premise attesting to the reliability of the original inference enhances the reliability of the conclusion of that inference, even though belief in the additional new premise is fallible. The answer lies in what the content of the additional premise does to the reliability of the conclusion. That premise tells us something *about* the inference, and that is crucial to the possibility of enhancing its strength. We can see this by noting that if the content of the reliability statement had been "Jones's is 50% reliable," rather than 90%, then we would not, intuitively, think that the added premise enhanced our justification for believing p. A favorable content in the added premise is not sufficient for the added premise to increase the reliability of the conclusion: if the belief in the statement "Jones is 90%" reliable were not itself justified, then it obvious wouldn't help us. But it is necessary to the enhancement of the original inference that the statement about its reliability level be favorable.

There are more sources of potential error in the new inference than in the initial inference, since the imperfect reliability rate of the belief that Jones is 90% reliable, say .95, combines with the original imperfect reliability rate of the belief that Jones said p, say, .95, and the content saying that Jones is 90% reliable introduces another 10% potential error. It follows from what we saw above that whether the parts are conjuncts or parts of a composition, the rule for reliability of the conclusion is to take the product of the reliability rates of the parts. In this case it will be (.95) x (.95) x (.9), or approximately .81, not at all shabby. The original inference had only one premise and one inference, but since on the matter of Jones's reliability the premise has no information, we must take that reliability as 50%. Then the corresponding numbers would be: (.50) x (.95) = .475. Thus, adding a justified belief in Jones's reliability did enhance the reliability and justification of the conclusion, despite the fallibility of every part of the proceedings, and the fact that there was an addition.

We get further confirmation for this picture of what is going on if we imagine a case referred to intuitively above, where the added judgment that Jones is 90% reliable is itself not 95% reliable, but as bad as possible: 50%. Now the calculation is (.95) x (.5) x (.9), approximately .43. Adding a completely unreliable premise only made things worse, despite the fact that the premise reported the new content that Jones is 90% reliable. The fact that adding fallible premises whose content does not enhance the support relation in the inference should make us less justified also falls out of this picture: If as we're assuming, the judgment that Jones said p is 95% reliable, then even if the judgment about Jones's reliability was 95% reliable, if the content of that judgment was merely that she's 50% reliable, then our calculation will be: (.95) x (.95) x (.50), or approximately .45.

Checking steps in a demonstrative proof has a similar structure. In checking such an inference we are forming a fallibly justified belief that the premise implies the conclusion. The fallibility of that belief increases the total potential error of the step in the proof, however slightly, since it is strictly an additional part. However, the content of the judgment about the inference is that it is a case of implication, i.e., that in making the inference one cannot go wrong. Compare now the initial inference, from A to B, with the inference augmented by this additional justified premise. In inferring from A to B, we had, say, .95 reliability or justification for the belief in A, and, say, .90 reliability or justification in our inference to B. Thus, the total error was (.95) x (.90), approximately .86. In the checked inference we have the same reliability of premise A, .95, and we have another premise, about the inference, with reliability of, say, .95, which will add potential error, but in addition the new premise says that the inference from A to B is 100% reliable. Thus, we have (.95) x (.95) x 1, approximately .90. Understanding proof checking in this way shows that despite the fact that potential error grows with the number of steps, intuition is right to think that the extra fallible step of proof-checking can increase our justification. Notice that this view does not take a stand on whether proof-checking is necessary in order to be justified at all in a step of proof; it simply shows how any justification level you have can be improved. This includes the case where you start with no justification at all.

Notice that in the case of demonstrative proof, it is natural for the premise that results from the check to say that the inference is 100% reliable, since the judgment would naturally be that this was a case of implication, rather than failing to be an implication. For this reason, the factor that we took as 1 above would be 1 in any normal case. This has interesting implications. For example, if you were so blessed that your original inference was infallible, adding a check of the inference would, on this view, necessarily reduce your reliability and justification, since the judgment "this is an implication" cannot improve on the infallible inference itself, and the addition of a fallible premise to the inference would still add potential error. It isn't just that if we were perfect we would not need to check our proofs, but that if we were perfect then checking our proofs would make us imperfect. But this is a special case of a more general point. It follows from the way that reliability and error accumulate in both the case of judging the reliability of the source and the case of checking a proof step, that if the checking premise you can justify attributes no more reliability to the inference than you already have justified, then you must not incorporate it on pain of strictly increasing your potential error. Such a premise will bring error, and nothing else with it.

Since the added premise in the case of checking a demonstrative inference is naturally the claim that it is a case of implication—that is, that the original inference indeed is 100% justified—the checking step can't increase reliability or justification more by claiming more. In such a case, the limiting factors on the ability of a proof-checking step to increase your reliability or justification level are 1) the potential error of the added premise about the inference, and 2) the level of reliability of the original inference. If the latter is high, then the added premise may not be able to do any good to your justification level, and may do harm if the fallibility of the added premise is high. This implies that it is salutary to be good at determining when a check is needed and when it is not, not just for pragmatic reasons of economy but for epistemic ones.

In the case of strengthening the reliability of your conclusion by adding empirical evidence to your premises, the question is again how this can happen when each added premise brings more potential error. The answer must be, as just discussed for the other cases, that when the reliability of the conclusion is successfully enhanced, something else is happening at the same time as the addition of premises. In the cases above, what was happening was that the content of the added claim was that the inference was so and so reliable. This is not something we have in the empirical case. Making this case as mundane, and clichéd, as possible, suppose we conclude that all swans are white on the basis of increasing evidence, more swans, all white, more swans, etc. The kind of evidence added is statements like “That swan is white,” wherein no mention is made of whether the inference from “These swans are white” to “All swans are white,” was reliable. However, as a matter of fact, more white swans do confirm “All swans are white” more.[iv] So, securing a reliable belief in that is securing a reliable belief in something that as a matter of fact enhances the support of the conclusion. It does not have to say that the inference is so and so reliable to make the inference be more reliable. If you were an internalist the way of telling this story would likely be slightly different. One would want to say that the inference’s justification is enhanced not just by the fact that another white swan has been found but rather by that in conjunction with the subject’s knowing that this new piece of evidence enhances the inference. Either way, it is now clear how it is possible for additional evidence to enhance the reliability of our conclusion despite the fallibility of the additions.

What counts as a step?

A given content can be distributed over beliefs in many ways. For example, “It is a grey split-level house” has a content that could also be expressed by three beliefs: “It is grey,” “It is split-level,” and “It is a house.” The more content there is, the more ways of distributing it there will be. I have claimed that one’s level of justification for a belief depends on how many steps of inference and how many premises one uses, among other things such as the error rate in each. But premises are beliefs, and how do we count them given that the same information can come in different packages? If individuation of steps and premises is an arbitrary matter, then one can count oneself to be as justified as one wishes.

What will the rule be for individuating beliefs over content? The answer, for our purposes, is straightforward. Although content *can* come in many different belief-packages, in a given inference from premises to conclusion it only comes one way. The premises of a given inference are identified by the psychological fact that the person inferring deployed certain of his beliefs, and not others, in coming to his conclusion—the fact that certain beliefs were the basis of his conclusion—and among those, we count the ones that affect the reliability of his concluding belief. The property we are judging, reliability, or justification, is not a property of a content, as the property would be if we were judging, say, the simplicity of the content “It is a grey, split-level house.” The fact that this content could be divided into three predications raises the question whether the content is to be counted as one (simple) or three (less simple). If we are asking about simplicity as a property of the content, it is inappropriate to regard it as determined by which way a given person actually believes the content. In contrast, in our situation the property in question is a property of a person's relation to that content via his beliefs, namely, whether or not his belief reliably reflects whether the content is true. Therefore it is perfectly fitting for the answer to be relative to what the beliefs he used in the inference actually were.

It is important that the beliefs at issue be not only possessed by the subject but also the ones he actually used in the inference in question. He may believe it is a grey house, but if he only uses his belief that it is a house, then the belief merely that it is a house is the one whose reliability matters to the calculation of the potential error of his conclusion. What if he uses more than he needs to? Suppose he uses his belief that it is a grey house as a premise, but only the part about its being a house is relevant to the conclusion he draws. Which belief's reliability rate do we figure in? This depends on what we mean by "using a belief as a premise." We might mean that he writes it down as part of the notes he scribbles while thinking, or says it aloud to himself while listing his reasons. These are not sufficient, though, to be taken as having used the greyness in his inference, since these actions are not sufficient to make his inference *depend* on the belief in greyness. He might have written or said these things while not having any disposition to withdraw his belief in his conclusion were he to find out that the house is yellow rather than grey, for example. There is more than one way to define dependence, and I provide the dispositional one only as illustration, but to have its error rate counted a belief has to be one the person's inference depends on, however exactly one will define that notion.

Whether dependence is sufficient for a belief's error rate to be included in the total potential error of his conclusion depends on what notion of reliability we are concerned with. For example, if we are concerned only with the rate of his conclusion belief being true when he believes it, then we are concerned with false positives. If our man behaves as if his conclusion depends on greyness, when it only depends on houseness, he will never believe the conclusion because of being wrong about greyness. If he doesn't believe in the greyness when it is there, then that will prevent him from drawing the conclusion belief. But that is not a false positive. If he believes in the greyness when it isn't there, then there are only two possibilities about houseness that could be part of a false positive belief in the conclusion: either he believes it's a house and it isn't, or he believes it's a house and it is. If the latter, then his conclusion will be positive but not false. If he believes it's a house and it isn't, then he would have produced a false positive conclusion without the help of the false greyness belief. If our notion of reliability is only concerned with false positives, then if the subject's inference depends on beliefs with more information than is relevant to the conclusion, his error rate for the part that is relevant is the appropriate one to use.

Things are different if we are concerned also with false negatives, and the case just described illustrates this. For the subject could have a true belief that it is a house, which would have led to the conclusion belief but didn't, because the subject felt the greyness was relevant and falsely believed that it wasn't grey. A false negative about greyness would produce a false negative in the conclusion. Thus, if we are concerned about false negatives as well as false positives, it seems we must figure the total potential error in the inference by using the error rate of all parts of the beliefs that the inference actually depended on.

One might worry that it is arbitrary that if the person uses the premise “It is a grey house” we might count him as having less potential error than he had had if his belief had been divided into two as in “It is grey,” and “It is a house.” However, that will not automatically happen. It would seem natural for a person to have as much chance of being wrong in his beliefs that it is grey and it is a house as he has of being wrong in his belief that it is a grey house. However, there are various ways this might not be. The most obvious is if he is unable to recognize the truth of “It is grey” and “It is a house” because the two properties are separate in these sentences, and he can only recognize the properties together. Granted, this would be unfortunate, but it illustrates the fact that whether someone believes something in a given situation—which is something the reliability rate strongly depends on—is in part a matter of his psychology. There isn’t an error rate we *should* attribute to the person for his belief that it is a grey house, given his error rates in believing that it is grey and that it is a house. Our questions dictate that we use the rates of error he actually has for those beliefs that he uses in his inference, which in principle may have any sort of relation to one another.

Conclusion

It is a familiar fact that error builds up over the course of inferences, and it is sometimes greeted with dismay and even avoidance. The fact that it builds up over strings involving not just conjunctions but also compositions of sources and beliefs about reliability has been less discussed. I have identified the rules by which potential error builds over such steps, and argued that the rate at which it builds is not out of keeping with how justified we typically feel we have a right to be in such situations. However, the fact that error strictly grows over steps does seem to undermine the possibility that reliability- and proof-checking make our conclusions more secure, and that increasing our empirical evidence for a claim makes our conclusions more justified. I have explained how these cases work in the detail required to see why proof-checking and evidence gathering are the boons we think they are.

References

Michael Detlefsen and Mark Luker (1980), “The Four-color Theorem and Mathematical Proof” *The Journal of Philosophy* LXXVII: 803–820.

Hume, David (1978), *A Treatise of the Human Nature*, Book I, Part IV, Section I; L. A. Selby-Bigge edited, second edition, with revisions by P. H. Nidditch, p. 180.

Jeshion, Robin (1998), *Philosophical Studies* **92:** 85–112.

Kitcher, Philip (1984), *The Nature of Mathematical Knowledge*. Oxford: Oxford University Press.

Michael D. Resnik (1989), “Computation and Mathematical Empiricism” *Philosophical Topics* XVII: 129–143.

Vogel, Jonathan (2006), “Externalism Resisted,” *Philosophical Studies* 131: 729-742.

[i] Of course, we need not assume this in order for the argument to work. I point out that we can assume this is obvious to you in order to highlight the fact the arguments I am making about justification should work whether we are internalists or externalists.

[ii] I take randomness rather than anti-reliability as the lowest level of reliability (0) because anti-reliability perfectly conveys information. 50% potential error, by contrast, conveys no information no matter how you slice it.

[iii] This is where the true instance of the Connection Thesis we saw above shows up, though of course here we are dealing with a more general phenomenon since here the belief claiming reliability of the source may not be justified. One might construe those as instances of the CT that are true in a degenerate way, but they are not instances that make the thesis interesting.

[iv] That is, provided they are a representative sample.