

# *Justification and the growth of error*

**Sherrilyn Roush**

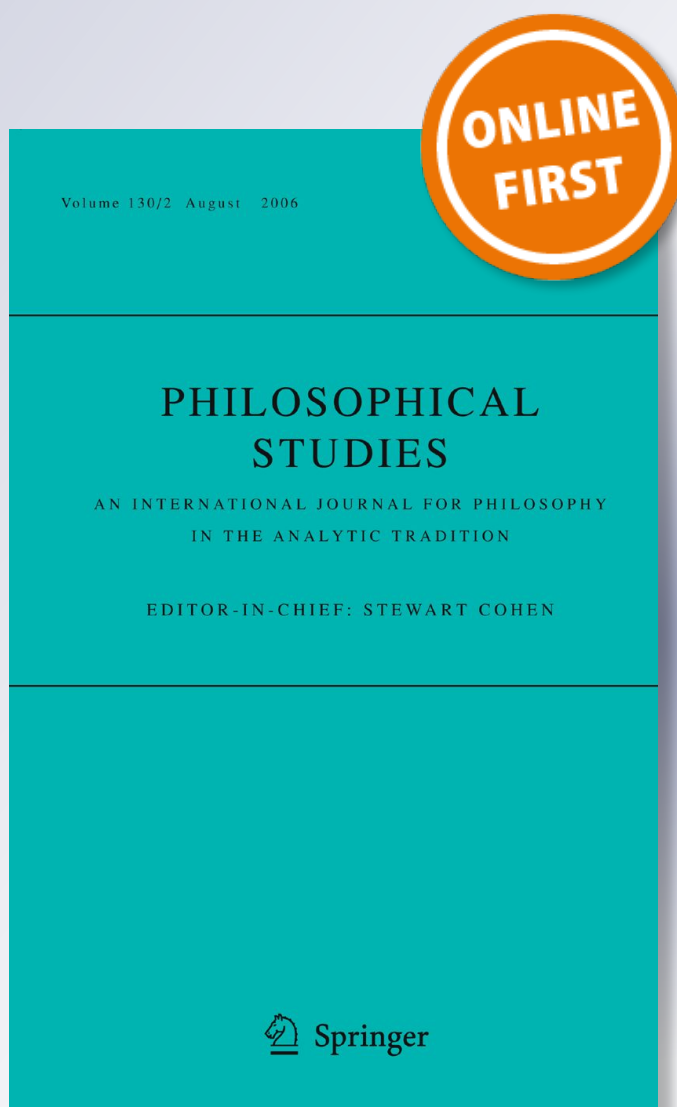
**Philosophical Studies**

An International Journal for Philosophy  
in the Analytic Tradition

ISSN 0031-8116

Philos Stud

DOI 10.1007/s11098-012-9967-7



**Your article is protected by copyright and all rights are held exclusively by Springer Science+Business Media B.V.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.**

## Justification and the growth of error

Sherrilyn Roush

© Springer Science+Business Media B.V. 2012

**Abstract** It is widely thought that in fallible reasoning potential error necessarily increases with every additional step, whether inferences or premises, in the same way that the probability of a lengthening conjunction shrinks. However, this has the absurd consequence that consulting an expert, proof-checking, filling gaps in proofs, and gathering more evidence for a given conclusion necessarily make us worse off, since they also add more steps. I will argue that the self-help steps listed here are of a distinctive type, involving composition rather than conjunction. Error grows differently over composition than over conjunction, I argue, and this dissolves the apparent paradox.

**Keyword** Justification · Growth of error · Internalism · Externalism · Connection thesis · Consulting experts · Reliability · Proof-checking · Reliability-checking · Double-checking · Evidence · Gap-free proof · Total probability

It is widely accepted that in fallible reasoning potential error necessarily increases with every additional step, whether inferences or premises, because it grows in the same way that the probability of a lengthening conjunction shrinks. As it stands, this is disappointing but not unintuitive. However, consulting an expert, proof-checking, filling gaps in proofs, and gathering more evidence for a given conclusion also add more steps, and we think these actions have the potential to improve our reliability or justifiedness. Thus, the received wisdom about the growth of error implies a skepticism about the possibility of improving our reliability through effort. Paradoxically, and even more implausibly, taking steps to decrease your potential error necessarily increases it. I will argue that the self-help steps listed here are of a

---

S. Roush (✉)  
Department of Philosophy, University of California, Berkeley, CA, USA  
e-mail: roush@berkeley.edu

distinctive type, involving composition rather than conjunction. Error grows differently over composition than over conjunction, I argue, and understanding how it does grow dissolves the apparent paradox.

## 1 The connection thesis (CT)

It is intuitively compelling that 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. However, this CT is apparently not true on any externalist view of justification (Vogel 2006). I will argue, first, that the CT 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. Its interest rather lies in the distinctive questions and apparent paradoxes about fallible inference and the growth of error over steps that it illustrates, as I will explain below.

The CT can be represented:

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

If I am justified in believing that 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 (or process), a source that produces true beliefs most of the time:

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

Thus for reliabilism to imply the CT 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  $q$ :

$$R(q) \rightarrow q$$

No one I am aware of defends an externalist theory on which having a reliable belief<sup>1</sup> that  $q$  implies that  $q$  is true, so  $R(R(p)) \rightarrow R(p)$  is not generally true either. It is tempting to think of this as a consequence of externalism itself. That one has a reliable belief that  $q$  is a fact about the world with a contingent relationship to the truth of  $q$ . If justifiedness is reliability, then justifiedness is automatically affected by this contingency. Thus exactly as  $R(q)$  comes up short when measured against  $q$ ,  $R(R(p))$  comes up short when measured against  $R(p)$ .

By contrast, for the intuitive notion of justification reliability claims appear to have a mysterious transparency that the CT brings out.  $J(R(p))$  does not imply  $R(p)$ ,

<sup>1</sup> I will often use the shorter phrase “reliable belief” to mean reliably acquired belief, where the upshot is that the belief is likely to be true because of its provenance.

that the source is in fact reliable, but on an internalist conception of justifiedness, the acquisition of belief in  $p$  actually *being* reliable does not matter to the belief in  $p$  being justified. This would explain the intuitive observation: if I am justified in believing the reliability claim about the source then that justifies me in believing what the source says.

However, it is natural to suspect that the failure of the CT for externalist theories is a consequence of fallibilism alone, the fact that the reliabilist view takes it as possible for a subject to be justified and wrong. Fallibilism is definitely assumed in the argument that the CT 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. And in the argument above that the CT is false for reliabilism, the mere assumption that reliability is fallible implies that  $R(q)$  does not imply  $q$ , which is the crucial step. Thus, fallibility of reliabilist justification is sufficient for that conclusion.

That the reliabilist notion of justification in question be fallibilist is also necessary for the conclusion that the CT is false for reliabilism. We can confirm this by considering the logical possibility of an infallibilist, externalist criterion for justification. On this view:

$$J(q) \text{ iff } R(q) \text{ and } R(q) \rightarrow q$$

If a belief is infallibly reliable, then the belief is true. Substituting for  $q$  the expression “ $R(p)$ ,” we get the following rendering of the CT:

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

which we now understand as saying 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  $p$ . Since the reliabilist equates justification with reliability, this implies that the CT:

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

is true. For an infallibilist reliabilist, the CT is true because perfect reliability is preserved under this composition.

If transparency follows from truth of the CT, then reliability is transparent to externalist justification as much as it is to internalist, if we raise the reliabilist justification requirement to an infallibilist level. Since the CT 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.

## 2 Internalism rebuffed

We can see the CT as true for reliabilism if we imagine reliabilism as an infallibilist view. But most of us view intuitive justification as fallible just like standard reliabilist justification is. Can we also see the CT as false for intuitive justification by focusing on the fallibility of that kind of justification, now going back to the

more natural assumption, for  $R(p)$ , 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  $p$ , that I needed a new battery. To see that the CT is false for internalism it is sufficient to ask how reliable I think the mechanic is, how justified we are assuming I am in believing that, and how my final justification for believing  $p$  varies with the answers to these questions. We may ask these questions two ways, one assuming that reliability and intuitive justifiedness are threshold notions, and one assuming that these properties come in levels; we are not justified or unjustified but justified to a certain degree and infallibly justified when that degree is 1. In the first way, I am asking how probable I think it is that this mechanic is right when he says  $p$  and how probable his being right needs to be in order for me to count him as *reliable*. About justifiedness I am asking how justified I need to be in order to count as outright *justified*.

In the second way of formulating our questions I am asking how reliable I believe the mechanic to be, and the “R” in our formulas will be replaced with “R-75” or R-95” as the case may be, depending on how often I think he gets these things right. I am asking also how justified I am in believing this about the mechanic, in the form of asking whether a “J-70” or perhaps a “J-85” is appropriate. I am considering these two ways of formulating the question whether the CT is true for internalism in order to show that my conclusion that it is not true does not depend on acceptance or rejection of a threshold notion of justification but only on the assumption of fallibilism.

Now consider a case in which the threshold for counting a belief as *reliable* (R) is when the source (or process) has an 80 % probability of getting these things right, and the threshold for counting someone’s belief as *justified* (J) is when it is at least 85 % along the way to deductively proved (however that might be understood: strength of grounds, number of reasons, etc.). Compare two cases, one in which I judge that the probability the mechanic is right is 81 % and the other that it is 99 %. Thus, in both cases I judge him to be something that counts as *reliable* (R). Suppose also that in both cases my beliefs about the mechanic are 85 % along the way to being deductively proved so that, by the set threshold, I am *justified* (J). Thus in both cases I count as *justified* in believing the mechanic is *reliable*. That is,

$$J(R(p))$$

For the CT to be true it must be that I am also justified in believing  $p$  in both cases. That is, for both cases it must be that

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

In order for me to be justified (J) in believing  $p$ ,  $J(R(p))$  must have given me 85 % justifiedness in believing  $p$ . Am I 85 % justified in believing  $p$  when I am, as in the first case, 85 % justified in believing the mechanic is 81 % likely to be right about  $p$ ? That is a bit uncomfortable. A sharper way to see the problem is that with these threshold notions I or my belief possess a status of justification (*justified*) with respect to  $p$  that is the same in the two cases I described, one with a judgment of 81 % reliability, the other with 99 %. If the two cases had involved two different

mechanics with what I (*justifiedly*) judged as an 81 % and 99 % chance of being right, then since they both count as *reliable* by my lights if CT is true I should be indifferent between them if one mechanic said  $p$  and the other  $\neg p$ . In fact, if CT is true, then I would be *justified* in believing *both*  $p$  and  $\neg p$ .

Similar difficulties arise when we avoid thresholds and talk only of degrees of justification and reliability. We imagine a set of cases analogous to that above. In the first I judge that the probability the mechanic is right is 81 % and the other that it is 99 %. Suppose that in both cases I am 85 % justified in believing these things. That is,

$$J_{85}(R_{81}(p))$$

$$J_{85}(R_{99}(p))$$

The CT's being true requires that:

$$J_{85}(R_{81}(p)) \rightarrow J_{85}(p)$$

$$J_{85}(R_{99}(p)) \rightarrow J_{85}(p)$$

It is hard to swallow that I should have the same level of justification for believing  $p$  when I trust a mechanic whom I believe to be 81 % reliable as when I believe him to be 99 % reliable. But even worse, taking ourselves as we did above to be considering two different mechanics and replacing  $p$  with  $\neg p$  in the second formula, we have the CT telling us that we are 85 % justified in believing  $p$  and 85 % justified in believing  $\neg p$ . We are equally justified in believing either  $p$  or  $\neg p$  despite the fact that we believe the source who said  $\neg p$  is almost 20 % more reliable than the other source. What this illustrates is the implausibility of the claim that intuitive justifiedness of belief is transparent to reliability claims.

The CT is not generally true, even intuitively, for fallible justification and reliability. 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 some grounds for justification are stronger than others. Moreover, the CT is false whether or not we adopt a threshold conception of justifiedness.<sup>2</sup>

The assumption that yielded this type of counterexample was fallibilism for intuitive justification and reliability-justification. If we assume that intuitive justification and (judged) reliability of the source are infallible, then the CT 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 the reliability I attribute 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.

This argument works the same if "J" refers to a fallible sort of intuitive justification, as long as the reliability of the source is assumed to be infallible. The believed infallibility of the "R" property means that the level of justification

<sup>2</sup> It is important for what comes below to note that this argument that the CT is false does not show that justifiedness always decreases when we interpose sources, even fallible ones. This is the easiest type of counterexample to see, but other types are discussed below.

indicated by the first “J” remains in the second “J,” whatever that level may be, since the R claim introduces no possibility of error or variation. However, the fact that the CT can be made true in this latter way, that allows fallibility of intuitive justification, does not help the internalist against the externalist, since the CT only turns out true here because the externalist’s criterion, reliability, is assumed to be *infallibly* fulfilled. If we assume that standard, then the CT is true for externalism too, as we saw above. Intuitive justifiedness is transparent to reliability claims only if they are claims of *infallible* reliability so the transparency that can be had is not due to any non-reliabilist quality of the “J.”

We have seen that if both J and R are fallible, then the CT 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 CT 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 attributed to the source by the claim  $R(p)$  is 90 %. We want to know whether:

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

The first J here is assumed infallible. Thus,

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

$R(p)$  does not imply  $p$ , since  $R$  is fallible. Since  $J(p)$  is infallibilist this link through a belief in  $R(p)$  whose level the subject has no reason to believe is infallible will not help get the subject to an infallible  $J(p)$ . And how else could  $J(p)$  follow from  $J(R(p))$  in the described situation except via  $R(p)$ ? Even if the intuitive justification of the belief that there is a (fallibly) reliable source is infallible, the CT is false.

Cases where the CT is false are all and only the ones where we have assumed the reliability claim  $R()$  to be fallible. Accepting this assumption, the CT is not true even if justification,  $J()$ , is internalist and infallible. Vogel is aware that what makes the CT 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 failure of the CT is apparently an expression of the familiar fact that justification diminishes the more steps we take because of the potential error in each one. However, the kind of step involved here is crucially different from those usually considered, and it leads to unfamiliar paradoxes, and opportunities, as I will discuss below.

The generic implication that potential error grows with steps is not novel, but the source of the phenomenon is often taken to be a threshold conception of justification: enough fallible steps, each with non-zero fallibility, will put one’s total potential error below the threshold of error-avoidance that was required for justification. However, as we have seen, the thesis is also false merely by the level of justification dropping by degree over steps. What mattered in the arguments above was not that the numbers associated with the first “J()” and second “J()” were one greater than and the other less than a magic number, but merely that the numbers associated with strength of justification could not plausibly be the same for the two “J()s.” Thus, the source of the failure of the CT is rather the conception of



justification, or the stuff underlying justification, as having comparative strengths at all, which it will need to for any fallibilist view.

If the CT is wrong, it is fair to wonder why it looked so right for intuitive justification. This, I think, is for at least three reasons. A suggestion that I will not try to develop in detail here is that the intuitive usage of the term “justified” is assuming a current-time-slice view of justification, namely, that the justificatory status of a belief depends only on properties or resources that the subject has at the time of the belief (Goldman 2008, p. 341). This would seem to imply that once the subject has taken what the source says into account, and taken the plunge of believing that *p*, the properties of that source are not relevant any longer to whether belief in *p* is justified. Metaphorically, once the belief is formed, the believed level of reliability of the source is screened off.

Though I think some such assumption is behind the intuitive plausibility of the CT, a current-time-slice view is not sufficient for the thesis. A counterexample will show this. Consider a reliabilist view of justification where the kind of reliability in question is probabilistic sensitivity of a belief. On this account a belief is reliable, hence justified, if and only if the probability is high that the subject does not believe it given that it is false, which I will express as “S(*p*).” This view does not appeal to anything about the history of the belief in question, but only to its dispositional properties at the time we are evaluating the belief. If this is the notion of justification, then the CT,

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

becomes

$$S(S(p)) \rightarrow S(p)$$

which is false: substituting *q* for S(*p*), this would require that  $S(q) \rightarrow q$ , i.e. the infallibility of this sensitivity criterion. A current time-slice view may be behind intuitive ideas about the CT, but it is not sufficient to imply maintenance of justification level over steps, and so not sufficient for the CT.

A second factor at work in the plausibility of the thesis 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 CT for it is more easily derived. Another factor that enhances the effect of this is that the phenomenon that makes the thesis false is the possibility of attrition of justification over steps; this attrition is thought to increase with the number of steps and may only become noticeable when there are many. Since the CT involved only two steps and the initial justification of the reliability judgment as well as the reliability claimed in that judgment were high, the degree of attrition present in the examples that inspired the thesis was small. Since most everyday instances of inference like those in the CT would involve the minimum number of steps, it makes sense that ordinary usage of the word “justified” does not force one to observe this difference. But I doubt that we would be as confident about a case like “If I am justified in believing that the witness reliably reported that a reliable party said *p*, then I am justified in believing *p*.”

A further contribution to the intuitive plausibility of the CT is that it looks like 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 may 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. The drop in justification would seem to grow uniformly. Whatever the calculus might be, there does not seem to be any way that combining a 95 % level with a 90 % level will bring you to no degree of justification at all, or to a random level of reliability. This is obvious in the commonly discussed way in which error grows over an increasing number of claims. Because the growth of steps there is by conjunction, the reliability decreases via a multiplication rule, and so has a lower bound on its rate of growth. It is plausible that there are lower bounds whatever the rule might be here.

So far I have considered cases and aspects of the CT that make it seem that the resulting justifiedness of a belief in  $p$ ,  $J_f(p)$ , will be weaker than the initial justifiedness in believing that the source we trusted was reliable,  $J_i(R(p))$ . These counterexamples suffice to show that the CT is false but do not tell us how error does grow. The examples do not even rule out the existence of instances where the final justifiedness for believing  $p$  is stronger than the initial justifiedness in believing the source was reliable. It will turn out that this is a good thing. To determine the fate of the whole scope of types of CT examples we must consider the thesis and structurally similar judgments more closely.

### 3 Error per step by conjunction and composition rules

Growth of error over lengthening of conjunctions is well-understood because the probability of a conjunction is a simple function of the probabilities of its conjuncts. If the conjuncts are independent, then we multiply the probabilities of the conjuncts. If they are not independent then we multiply the probability of one conjunct by the probability of the other given the first; either way the answer is a product. If each conjunct might be wrong, or might be wrong given the other, then its probability, or probability given the other, is below 1. Thus we are multiplying fractions of 1 and each new one lowers the probability of the conjunction. For this reason when additional fallible beliefs are conjoined to existing ones they strictly lower the reliability of the conjunction of beliefs, thus lowering one's externalist justification for believing the set of them and thence, it seems, for any conclusion inferred from them. Since this can be obvious to a subject, internalist justification involves this attrition as well. There can also be a growth of error without an increase of premises, in a proof to which we add more and more inferences; since each added inference adds another source of potential error this potential also appears to accumulate by conjunction. Therefore, the more the premises, and the longer the proof, the greater the potential error.<sup>3</sup>

<sup>3</sup> Here and in what follows we will appear to be counting steps, e.g., claims, inferences, checks, in order to assess how much potential error is generated. One might wonder what individuates the steps. What if a person has a belief that it is a gray house? Does this count as one claim—that it is a gray house—or two—that it is a house and that it is a gray house? The answer is straightforward. It is one or two depending on whether it contributes as two beliefs or one in the actual inferences, beliefs, and checking procedures of

The conjunction model of the direction and pattern of this phenomenon accords with our intuitions in many straightforward cases. When we seek information we try to get as close to the source as feasible since fidelity potentially drops off with every middleman. The reason is exhibited in the game Telephone where a message is whispered from one person to the next along a line of, say, a dozen people; we are not surprised when the message coming out the end of this line bears little relation to the message that went in. The law is strict about sticking with the original source; 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. This is at least partly because more links increase potential error and the quality of the link the hearsayer is using is not something the jury has the means to judge directly in such a case.

Potential error seems to increase over links, and how much it does depends on the quality of the links, as the conjunction rule predicts. We trust a report from a science journalist if we have no access to the scientist, exactly to the extent that we judge the journalist a reliable transmitter. We tend to trust that a photograph of a painting that presents itself as accurate looks just like, or largely 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 book dealer. But when we do this it is because the photo comes through a reputable museum or publishing house. Photographs are faithful to a high degree unless doctored, and the institutions in question have high incentives not to doctor at any step—such things are easily checked, and a good reputation is hard to regain once it is lost. At least when we do trust the photograph, these are the kinds of things we are presuming about the intermediate links.

In familiar cases the conjunction rule model seems to get the direction of error growth right, and the fact that its rate depends on the quality of each link, but the rate of error growth appears to be too high to make sense of some examples, such as multi-million-dollar art auctions. There are conceivably many links in the chain of events taking an artwork from the owner to the auction: truck drivers, guards, technical people, administrative assistants, dealers. If there are fifteen links in this chain, and 99 % fidelity at each link, then fidelity at the end is only guaranteed at 86 %. It seems unlikely that a buyer would bet millions of dollars on an 86 % chance the painting is authentic. Perhaps each link is more reliable than I estimate—if it is 99.9 % then the resulting reliability is at least 98.5 %—but can the fidelity of each link really be that high? Perhaps that doubt is why a buyer typically brings his or her own expert to verify authenticity. However, this brings the conjunction rule model to a serious problem: consulting an expert is a further step, and each additional step is supposed, necessarily, to increase the potential error. How can consulting an expert make the buyer's predicament any better?

We are accustomed to the fact that more middlemen, more inferences, and more claims introduce a greater chance of error. However, we also expect that in each

---

Footnote 3 continued

the subject in the argument in question. How we count thus depends on how one defines what it is for a belief that a person possesses to be a *basis* for inference, checking, or other beliefs, but we do not need to take a stand on that question here.

case it is possible to improve a given situation by applying more effort to it. We expect, for example, that it is at least possible for checking with a mechanic to make us more reliable and more justified in believing we need a new battery than the reliability or justification we could have had without adding that check. It is similar with proof-checking. It may or may not be strictly required for justification in mathematics (Detlefsen et al. 1980; Kitcher 1984; Resnik 1989; Jeshion 1998) but it should at least be possible for checking our inferences to enhance our reliability or justification. Similarly, even though additional evidence increases the number of premises in our argument it should at least be possible for more evidence to increase our reliability and justifiedness in drawing a given conclusion. Finally, if constructing the gap-free proofs that Frege proposed has no other virtues everyone seems to agree that if we did them we would, at least under some conditions, be epistemically more secure in drawing our conclusions. After all, the smaller the steps the less likely in each one of them that we will make a mistake.

However, checking with a mechanic and checking our inferences make for more steps, taking more evidence on board adds premises, and filling the gaps in proofs increases the number of inferences. Each step added, whether by check with an expert, a new premise, acquiring new evidence, check of an inference, or interpolation of an additional line of proof contributes a new source of potential error; if these combine by the conjunction rule then a greater number of fallible steps cannot even allow us to maintain a level of justification or reliability, much less to improve. The apparent paradox in this can be put this way: in these cases you appear to be necessarily *less* justified though you have taken quite sound and we think effective steps to make your conclusions more secure. Indeed, you have taken the only steps it seems we know of that could possibly help.

In this section I will argue that error depends on the number of steps in argumentation differently over composition of reliability claims, than the familiar way error depends on the number of conjuncts in a conjunction. Below I will argue that understanding these differences resolves all of the versions of the apparent paradox, such as with double-checking an argument, increasing evidence for a single conclusion, and interpolation that fills gaps in proofs. To begin, consider again my use of the testimony of the mechanic, call him Terell, to form a belief that *p*, I need a new battery. I judge him 90 % reliable, and I am reliable to extent .8 in so judging, we imagine. Intuitively, these two things make me to some extent justified (or reliable) in believing *p*. But, again intuitively, two such reliability facts do not generally make me exactly as justified in believing *p* as I was in believing the mechanic was reliable. How justified do such facts make me?

The answer is that it depends. To explain this, I will interpret justification as reliability and simplistically express reliability of a belief in *p* as probability of *p*. Potential error is then the deviation of this probability from 1. This represents reliability via the upshot it has for our level of justification because I will interpret that probability as rational degree of belief of a subject. Interpreting the probability objectively will not affect my points as long as it is done consistently, and analogous arguments could be made for internalist justification composed with a reliability judgment since the points I will make can also be judged by a subject herself. This

picture is simple, but the model that results is sufficient to bring out structure that the conjunction rule misconstrues, and solve the puzzle above.

The key is that the subject's justification in her belief about Terell's reliability and what she judges his reliability to be do not combine as conjuncts when they go to give the subject her degree of belief that she needs a new battery. As the syntax in the discussion above of the CT showed, it is a relation of composition: the subject is justified in believing *that* Terell is reliable, and so, sometimes, justified in believing *p*. Crucially, when she asserts her conclusion *p* she is not asserting the conjunction of the claims that Terell is so and so reliable and that she believes that. She is, as it were, using these beliefs of hers rather than asserting them. Strictly, composition should be modeled with second-order claims, but fortunately that is unnecessary, since the relevant dependence between the composed parts can be brought out and faithfully understood using simple conditional probabilities.

The key is to understand what question we are asking when we want to know the total potential error that has accumulated over the course of our steps. We are not asking how much potential error there is in a conjunction of statements corresponding to those steps. What we want to know is how much potential error is in the conclusion-belief. In the case of the mechanic we are asking for the probability of *p*, that I need a new battery.<sup>4</sup>

If I form a belief that I need a new battery by consulting a mechanic I regard as reliable, then on the way from me to my having a belief that I need a new battery, there are two possibility forks: I am either right or wrong about the reliability of the mechanic, and he is either right or wrong about *p*. There is a probability that I am right about Terell's reliability, .8, and a probability of *p* given that I am right. The latter is given by the reliability I attribute to Terell, since that attributed reliability is the rate at which I think he gets *p*-like things right, i.e. the probability of *p* given that he attests to it, and we are simply assuming that he attests to it. I think Terell is 90 % reliable about *p*, and either I am right to think this, in which case *p* has .9 probability, or I am wrong, in which case since as far as our assumptions go Terell's reliability could be anything other than .9, we must regard the chances of *p* as 50–50. The probability of *p* is an average of how likely *p* is when I am right about Terell's reliability and how likely *p* is when I am wrong about Terell, a term called the “catch-all”, weighted by how likely it is that I am indeed right and wrong about Terell, respectively. That is, we use the rule of total probability:

$$\Pr(p) = \Pr(p/R)\Pr(R) + \Pr(p/-R)\Pr(-R)$$

where *R* = Terell is .9 reliable, and we assume as background that Terell has attested to *p*. In this case we are considering, I am .8 likely to be right in judging Terell .9 reliable, so  $\Pr(R) = .8$ . The probability of *p* given that Terell is .9 reliable (and he has said *p*) is of course .9. Thus the first term in the equation is  $(.9)(.8) = .72$ . I have a 20 % chance of being wrong to think Terell is 90 % reliable about *p*, so without any further information we must say there is a 20 % probability that Terell's reliability,

<sup>4</sup> In this sense the problem is very different from the preface paradox in which the conclusion whose reliability or probability we are interested in is itself a *conjunction* of the premises. My discussions in this paper provide nothing helpful to those difficulties since the conjunction rule is the correct one for that question.

and hence the probability of  $p$ , is .5. This makes  $\Pr(-R) = .2$  and  $\Pr(p/-R) = .5$ , which makes the second term .1. The resulting probability of  $p$  is thus .82. It is less than what I judge is Terell's probability of being right about  $p$ —to use the notation earlier,  $J_f$  is less than  $R$ . However it is not as low as it would be by the conjunction rule: it is .82, whereas the conjunction rule would have said it was .72.

The fact that the structure of these nested claims is not that of conjunction is reflected in the fact that the right-hand side of this formula is not a product. The intermediate events between my judgment of Terell and my conclusion about  $p$  yield a disjunction, and therefore a sum. If we multiply two fractions each less than one the result will always be less than either of them but that does not apply to an *addition* of products, even when they are fractions of one. The second summand in the total probability equation will always be positive or zero, and the conjunction term that we would have had under the old picture is the first term, so the reliability that composed claims leave us with is always greater than or equal to that with the corresponding conjunction. The properties of the total probability equation provide reassurance about the growth of error that potential error need not grow as fast with the number of steps as we had assumed. This does not yet touch the question of whether further steps and consultation can reduce our potential error, in other words enhance our reliability over what it was without the additional steps, but it will.

The case of Terell provides a preview of how it does that. The first step is to see that not only did our error not grow as fast as feared, it is also possible for my level of reliability about  $p$  by means of a consultation with Terell to be higher than the level of reliability with which I judge him. That is, in our terminology above concerning the CT,

$$J_i(R(p)) \rightarrow J_f(p)$$

the final level of justification indicated by " $J_f(p)$ " can be greater than to " $J_i(R(p))$ ." We see this in the example just developed, where  $J_i(R(p))$  was .8, but  $J_f(p)$  was .82. In the total probability formula this corresponds to  $\Pr(p)$ , the rational degree of belief in  $p$  that I end with, being greater than  $\Pr(R)$ , my rational confidence that Terell is .9 reliable. One thing that makes this possible is that unlike  $\Pr(R)$  and  $\Pr(-R)$ ,  $\Pr(p/R)$  and  $\Pr(p/-R)$  are independent of each other, and so may take any values relative to each other. " $\Pr(p/R)$ " is the probability of  $p$  when Terell is .9 reliable, which is .9. The key to the aspect that can yield fidelity beyond the conjunction rule is what the probability of  $p$  is, in other words what the reliability I attribute to Terell is, if it is not .9. What would his reliability be in the 20 % of cases where I was wrong about him?

With no information above, we assumed random, 50–50, (and still had  $J_f$  greater than  $J_i$ ) but we often have more information than that, information that would make the likeliest alternative reliability be more than chance, for example in cases where we are reasonably sure that his reliability lies within a certain range, say between .7 and .9. Suppose we have reason to be 80 % sure that he is 90 % reliable, and otherwise he is somewhere between .85 and .95. Thus, instead of  $\Pr(p/-R)$  being .5 we have it at between .85 and .95. Then .8 reliability that he is .9 reliable, with the just-stipulated .2 chance that it is between .85 and .95 instead, yields a resultant reliability of between .89 and .91 for the belief in  $p$  that I get by trusting Terell. This is much greater than the reliability I had in judging him,  $J_i$ . In this case the process of trusting Terell makes me more reliable than any of the judgments of mine that

went into that process. The general condition for trusting Terell to be a step that enhances my reliability in this way can be given in an equation.<sup>5</sup> One interesting upshot of it is that in the limit as  $\Pr(R)$  and  $\Pr(p/R)$  both approach 1 the enhancement described goes to zero. This accords with common sense, since if I am perfectly reliable in judging that Terell is perfectly reliable, then consulting him cannot add anything; since I have perfect reliability in judging his expertise, I must know as much as one can possibly know for being reliable about  $p$ .

What we need to address the self-help paradox, though, is not just that  $J_f$  can be greater than  $J_i$  in the situation where I did check with Terell to get my belief in  $p$ , but that what I get by *adding* the step of checking with Terell can be a better justification than what I could have had without adding the check. That is, it must be possible for the  $J_f(p)$  in

$$J_i(R(p)) \rightarrow J_f(p)$$

to be greater than

$$J_i(p)$$

But this easily follows from its being possible for  $J_f(p)$  to be greater than  $J_i(R(p))$  as just shown, and the fact that  $J_i(R(p))$  is the upper limit of the level of justification I could have had if I had gone straight to a belief in  $p$  without consulting Terell. This is because, in cases where we do think enhancement can occur, I must be a person who knows no more about  $p$  than the information that went into my judgment of Terell, and I probably know much less. Thus, the fact that the conclusion-belief whose error profile we are interested in is not a conjunction means it is possible for adding a check with a source we judge reliable to improve our epistemic situation with respect to  $p$  over what it was without that step, and the mechanism is sensible.

This mechanism is also the key to resolving the apparent paradox with proof-checking, or double-checking an argument, but to get there we first need to understand how the amount of potential error in the conclusion of an argument depends on the number of steps of the argument. We will find that a more general instance of the total probability model makes sense of the growth of error here as it did in the case of checking with an expert, and here as there it allows for greater reliability at a given number of steps than the simple conjunction rule did. For technical reasons having to do with the limitations of probabilistic analysis I will discuss arguments with defeasible inferences rather than logical implications, although I will sometimes slip into using the term “argument” interchangeably with “proof”; with care the points I make can be extended to the latter case.<sup>6</sup>

<sup>5</sup> If  $z$  is  $\Pr(R)$ ,  $y$  is  $\Pr(p/R)$ , and  $x$  represents  $\Pr(p/-R)$ , then  $\Pr(p) > \Pr(R)$ , i.e.,  $\Pr_j(q) > \Pr_i(q)$ , iff  $y + (x/z)(1 - z) > 1$ . Note that what is on the left hand side of the last inequality is not a probability; the formula is a relation between probabilities.

<sup>6</sup> The problem is that any inference from a claim to a claim that it logically implies must be given probability 1 or rational confidence 1, on pain of incoherence, simply because it is in fact an implication. In cases of logical and other necessary truths probability does not track epistemic legitimacy but the relations the propositions stand in as a matter of fact. Because of this, the fallibility of logical inference cannot even be represented on normal axiomatizations. However, there are re-axiomatizations that would allow the points here to be made for demonstrative proof because they allow logical propositions to be

Suppose we have an argument with four steps, premise A, intermediate conclusion from that, B, intermediate conclusion from that, C, and ultimate conclusion from that, D.<sup>7</sup> In making the argument we make claims A and B and C and finally D, each line inferred from the previous. When we finish the argument we do not assert the conjunction of D with the previous lines. We rather assert D and understand that how good that assertion is depends on the relations D (my rational degree of belief in D) has to previous lines (rational degrees of belief), directly and through intermediaries.

The total probability rule brings out these dependencies and tells us exactly, and intuitively, how the confidence in D that would be rational for us depends on our inferences between and D.

$$\begin{aligned} \Pr(D/A) = & \Pr(D/C)\Pr(C/B)\Pr(B/A) + \Pr(D/-C)\Pr(-C/B)\Pr(B/A) \\ & + \Pr(D/C)\Pr(C/-B)\Pr(-B/A) + \Pr(D/-C)\Pr(-C/-B)\Pr(-B/A) \end{aligned}$$

This incorporates all the forks and therefore possible errors in the path from A to D. In general, there will be  $2^n$  summands for every n claims in between A and D. This formula says that the probability of D given A depends on the extent to which C confers firmness on D as well as how likely it is that C is true, but also on how much firmness -C confers on D and the probability that C is false. I believe C in virtue of believing B and that link is being checked too, with the same weighted average of the possible ways that error could creep in. So,  $\Pr(D/A)$  also depends on how much firmness B and -B respectively confer on C and how likely that each of them is the case. This in turn depends on whether A confers firmness on B and whether A is true.

If the simple conjunction rule were used on this the probability of the conclusion of the argument given its premise would have been represented as  $\Pr(D.C.B/A)$ , and would equal the first summand of the right-hand side of the equation above. The existence of more summands in our equation confirms that it is possible for the probability of the conclusion to be higher than the simple rule implies, and it never will be lower. That addition is there because of the potential nodes of dependence nested between A and D, and those will be present in any argument where the subject has used steps involving beliefs in B and C; the (fallible) extra steps reduce the first term, but also provide the opportunity for more reliability via the additional summands.

One way to think of this is that belief in B and belief in C are intermediate “events” in the series of events beginning with belief in A and ending with belief in D. The same model works for consulting an expert source like Terrell and for figuring potential error in a proof because conditional probabilities in the proof, such as  $\Pr(C/B)$ , pertain to our inferences, and each is essentially an evaluation of

---

Footnote 6 continued  
given non-extreme probabilities (see Hacking 1967; Garber 1983. Garber’s view has the advantage of preserving coherence.

<sup>7</sup> If the argument has more than one premise, we imagine all of them loaded into A in the first line. The total probability model can be generalized to multiple premises, but it will have more terms because we are splitting the conjunction of all the premises into a list of conjuncts.



whether the previous line is a reliable source for the current claim—e.g., that the probability of C given B is high. Sources that are people can also be composed in long strings, and the same model applies. Indeed, the game Telephone described above is a series of nested sources, and also is not a case where the conclusion whose error profile interests us is a conjunction. Let

A ... Player 12 whispered q.

B ... Player 11 whispered q...

Z ... Player 1 whispered q,

where Player 12 is the one whispering in your ear. You want to know whether Z is true given that you believe A. Your confidence in A combined with lots of other premises will give you a confidence in Z, and those other premises are conditional probabilities. They are the probability that Player 11 whispered q given that Player 12 did and given that Player 12 did not, and so on back to the beginning. We thus have a probability of A computed by the total probability rule.

Cases properly described by the total probability rule can in principle have much lower total potential error per number of steps than those described by the simple conjunction rule, but, in accord with experience, Telephone is not a case where we should expect the difference to be very noticeable. This is not only because whispering is a low fidelity method, which affects the first term, but also because the extra terms that total probability has us adding will not be at all high in this case due to the nature of the infidelity. The extra terms have factors corresponding to the probability that player  $n$  whispered q given that player  $n + 1$  did not whisper q. If player  $n + 1$  did not whisper q then he probably whispered something that sounds somewhat like q, so something of that type is what we should think player  $n$  whispered in  $n + 1$ 's ear in that case. However, something that sounds like q but is not q is unlikely to mean the same or even something similar to what q means. As we saw above with Terell, we get the most out of those extra terms when the case where the source is wrong is one where it is not too far wrong, and in Telephone the lack of strong relation between phonemes and meanings prevents that possibility.

The cases of the museum photograph and the artwork for auction must also be rendered by total probability, since the final assertion whose potential error we are interested in—e.g., the photograph or painting is the real thing—is not a conjunction of assertions about the steps along the way. In these cases the links can have a quite high fidelity, making possible more reliable end products than in Telephone, and we tend to think it is possible for the final conclusion to be at least somewhat reliable. The total probability rule allows us to see how greater fidelity is possible per step than the conjunction rule said, but how much greater depends on the extra catch-all terms, which in this case, taking D as the claim that the last transfer, to the buyer, was of the real thing, involve the probabilities that transfer  $n + 1$  was of the real painting or photo given that transfer  $n$  was not. Unfortunately, those probabilities are negligible—it is difficult to transfer a painting or photograph that the previous party did not transfer to you—so we cannot expect to do appreciably better than the conjunction rule says we would.

It may be that we never really experience very high confidence in photographs, or that we are irrational if we do. But multi-million dollar art auctions do exist, and this result makes them puzzling due to the calculation we did above of how bad the conjunction rule verdict is. This may be why a big buyer hires an expert of her own to inspect the object she is buying for authenticity at the end of the chain of transfers. It is similar when we consult an expert about whether a particular photo was doctored. We think they have skills for detecting such fraud by direct inspection. However, these are extra steps in the process, so how can they improve the reliability of a belief about the end product?

We have just been comparing error growth per step for the conjunction rule and the total probability rule and found that total probability can in principle give us higher reliability in the conclusion than the conjunction rule led people to expect. Now we must ask whether according to the total probability rule your reliability can increase with addition of steps like forgery-checking, fraud-checking, double-checking an argument, filling gaps in an argument, and adding evidence to your premises. This enhancement through a further step was possible in the case of Terell, but that was less complex than these other cases. It does not seem possible to improve the reliability of the conclusion-belief in Telephone by adding another player unless he is a magician who knows what player 1 said independently of the chain of whispers, since magic seems to be the only way we could make the added catch-all terms greater than the initial conjunction terms. Are the photo expert and the art expert magicians? We need to look more carefully at what enhancement depends on.

#### 4 The growth and loss of error with additional steps

In all of the apparently paradoxical cases I have mentioned, the number of judgments we make along the way to the conclusion is greater than it was in our initial way of coming to the same conclusion. However, in many of these cases there is a special phenomenon that can render the accumulated potential error of the original process independent of the potential error in the new assertion of the conclusion. This phenomenon, which I will refer to as replacement, is nicely illustrated by the process of filling gaps in proofs.

Another feature that is needed for attrition even if the conjunction rule is the right one, or total probability is dominated by the term corresponding to the conjunction rule, is not present in the attempt to improve one's reliability by constructing a gap-free proof.<sup>8</sup> The new sequence of claims and inferences in a gap-free argument, though longer than that of the original argument, does not contain the old set as a subset. Because of this, even if error did grow by the conjunction rule it would be possible to improve one's rational confidence in the conclusion, because the probability of the longer set is independent of that of the shorter. If the gap-free argument becomes your psychological basis for believing the conclusion, the initial,

<sup>8</sup> The argument here works for defeasible arguments, and would need to be redrawn for demonstrative proofs in a probabilistic framework that could give non-extreme probabilities to necessary truths (see footnote 7).

shorter argument is epistemologically irrelevant to the potential error of your conclusion-belief.

To see these points, consider an argument from A to G:

A  
 D  
 F  
 Therefore, G

There are 4 claims in the argument, and we are imagining that each line's probability is dependent on that of the previous as in a proof—all premises are loaded into the first line—so there are 4 claims and 3 inferences. The subject believes A, infers D from it, infers F from that, and from F infers G. Making this argument gap-free would involve breaking down each of those inferences into a list of “smaller” intermediate inferences, so that the total number of claims and inferences would necessarily increase. For example:

A  
 B  
 C  
 D  
 E  
 F  
 Therefore, G

However, though the claims A, D, F and G, and the inference from F to G, all occur in the new proof, the original inferences from A directly to D and D directly to F do not. To strengthen performance at those links we *replaced* them. Thus, even if we viewed the set of beliefs relevant to the final probability of G as a conjunction of claims and claims about inferences, and so used a product rule, there would be nothing preventing the longer argument from yielding a higher probability for G; the probability the longer argument gives to G is not a multiple of the probability the shorter argument gives to G. The probability of a conjunction of claims, p, q, r, s, t,

$$\Pr(p.q.r.s.t) = \Pr(p/q)\Pr(q/r)\Pr(r/s)\Pr(s/t)\Pr(t)$$

may be higher than the probability of a shorter conjunction of claims, p, s', t,

$$\Pr(p.s'.t) = \Pr(p/s')\Pr(s'/t)\Pr(t)$$

because the place where these two conjunctions differ, at q.r.s in the one and s' in the other, bear no necessary relation to each other. The gap-free proof will contain all of the original *lines* as a subset of its lines, but not the original inferences to and from those lines and so not the same dependence relations. Gap-filling gives us a new proof so there is no surprise that it could have higher fidelity.

When the interpolated steps in the longer conjunction of the gap-free argument are collectively more secure than the single inference they replaced, the new argument will yield a higher probability for its conclusion given its premise than the original argument did. The smaller inferences in the gap-free argument will tend to be more secure if each of the new inferences is much easier to scan or comprehend

than the corresponding original inference, and they will be easier if they are “smaller.” This is how we think of the situation intuitively and it is to the extent that it is easier to avoid mistakes in these smaller inferences that we think filling in gaps is worthwhile.

We know that gap-filling does not make us more reliable in every case. For some people a shorter proof will be more reliable than a less gappy one, and our equations capture that too. Someone who was already infallible would not be helped by gap-filling—which is displayed in the total probability equation by the fact that an infallible reasoner already has reliability 1 for, for example, the inference from A to D, and no probabilistically-expressed reliability is going to be greater than 1. More realistically, we can imagine an especially talented or experienced reasoner for whom small steps would be less reliable than big ones—perhaps she is highly reliable in the big steps and the tedious bookkeeping of the gap-free proof, her lack of practice with it, and her conviction that it is unnecessary, make her more liable to err than she is with the big leaps. Or perhaps a person just gets too bored to concentrate or too confused by the similarities of the lines, or too lost in the symbols required to do an argument this formally, and so would be best off not trying it. In these cases we can imagine the error contributed by each of the greater number of lines as high enough that the total error grows; the greater number of steps, each one no more reliable than the gappy steps, would make the product of the corresponding probabilities smaller than it was for the gappy proof (and the extra terms that came with a proper total probability calculation would be low due to the lack of reliability of the added steps).

It is easy to see that there is replacement rather than supersetting in the case of Terell too. The question for the paradox was whether consulting Terell could make us more justified about  $p$  than we would have been without him; we know that it can in real life, and we saw above that it can according to total probability. It is indeed more steps if I both judge him and then from what he says I judge that  $p$  than if I go straight to a judgment of  $p$  without him. But the latter is not a subset of the moves I make in the former. That straight inference does not appear in the longer set of inferences via Terell, so the conjunctions of claims about their reliabilities bear no necessary relation to each other.

The case of accumulating more evidence is also one of replacement. In this case we are adding premises and so more sources of error, but what we come to at the end of an argument that appeals to evidence is not the assertion of the conjunction of every line or premise or inference that appeared in the argument; it is the assertion of the conclusion claim on the assumption of the premises. When we acquire a new confidence in our conclusion, we are not simply adding a premise—and new source of error—and repeating the same inference to the conclusion. The new inference is not the same as the old one because it is an inference *from* a different claim or set of claims, and it does not contain the original inference as a subset. Accordingly, adding a premise has the capacity to change the probability with which the total evidence set supports (reliably testifies to) the conclusion, so changes the strength of the inference, and can do this in a positive or negative direction.

An example will illustrate. Call our initial evidence set  $E$ —say observations of many white swans—and the new piece of evidence  $e$ —a sighting of another white

swan—and call our hypothesis  $h$ , that all swans are white. Let  $\text{Pr}_i$  be the initial probability function and  $\text{Pr}_f$  the final, after taking the new evidence into account. Our total probability calculation would tell us that:

$$\text{Pr}_f(h) = \text{Pr}_i(h/E.e)\text{Pr}_f(e/E) + \text{Pr}_i(h/E.-e)\text{Pr}_f(-e/E)$$

This happens also to be the formula for Jeffrey Conditionalization, which is the general learning rule in a probabilistic framework, the rule by which we update our whole set of beliefs on making a new observation. The final probability of  $h$  is given by an average of how likely the hypothesis is with the new evidence added and how likely it is without, weighted by how likely the new evidence is or is not. Because of the second summand, the catch-all,  $\text{Pr}_f(h)$  can be greater than  $\text{Pr}_i(h)$ .

The new evidence set is a superset of the old one, so the conjunction of its members does have a probability less than or equal to that of the old one's members, but this conjunction does not determine that the probability of the conclusion is lower on the new evidence set. To translate the current conclusion fully into the evidence context, in non-monotonic reasoning a new piece evidence can be a defeater and yield a lower conditional probability of the conclusion, or an enhancer and yield a higher conditional probability of the conclusion. Translated into this context the point is a truism.

There is a basic kind of modification of a proof or argument, perhaps the most common kind, that on the face of it does not involve replacement, which is lengthening a proof by adding steps at the end. The game of Telephone as it is played is also such a case. In the case of proof, the new proof is a superset of the original proof, both in the premises and in the inferences, because the most recent line and inference has been added without any other modification. Even if an argument is not a replacement for a previous argument but an extension of it, the total probability rule allows in principle that the longer argument be more secure than the conjunction rule says it would, as we saw above with Telephone and the art auction. But the conjunction rule verdict is always a loss, so for cases without replacement we need to ask whether the extra terms we get through the correct analysis more than compensate for it. Only so will we explain how enhancement of reliability of the conclusion can be achieved in a case where the number of steps strictly increases.

If enhancement and preservation of reliability over strict lengthening are not possible, then we have a serious problem, for we must wonder how we end up with the slightest trust in what mathematicians and scientists say, when their proofs and arguments are long and complicated, and pile inferences on the products of previous inferences. The total probability rule does allow the possibility of increasing reliability with a strictly increasing number of steps in principle, but the further question is whether the possible ways of avoiding growth of error are powerful enough and realistic enough to explain how science and mathematics are possible. Fortunately, I think we can also answer this in the affirmative.

Suppose we have a short argument from  $A$  to  $B$  and  $B$  to  $C$ , which we extend with a step from  $C$  to  $D$  to make a longer argument. The probabilities of the conclusions of each argument are:

$$\Pr(C) = \Pr(C/B)\Pr(B/A)\Pr(A) + \Pr(C/-B)\Pr(-B/A)\Pr(A)$$

$$\Pr(D) = \Pr(D/C)\Pr(C/B)\Pr(B/A)\Pr(A) + \Pr(D/-C)\Pr(-C/B)\Pr(B/A)\Pr(A) \\ + \Pr(D/C)\Pr(C/-B)\Pr(-B/A)\Pr(A) + \Pr(D/-C)\Pr(-C/-B)\Pr(-B/A)\Pr(A)$$

All of the terms in the calculation for the shorter proof (and some terms determined by them) occur in that for the longer proof, and the first thing to see is that though the longer proof does not involve replacement it has all of the features replacement gave that were needed for the enhancement of reliability. In the longer proof of D there are two new probabilities,  $P(D/C)$  and  $P(D/-C)$ , and two new summands (the second and fourth). Moreover the new terms  $P(D/C)$  and  $P(D/-C)$  on which  $P(D)$  depends are completely independent of the terms carried over from the previous proof, and of each other. At the level of description that matters to the possibility of maintaining and enhancing fidelity, replacement and lengthening are the same.

The two new summands are crucial to enhancement and preservation of reliability over lengthening. If we had only the first and third then in moving to the longer proof we could at best preserve the fidelity we had in the proof of C, and that would arise only if  $P(D/C)$  was 1, which is unrealistic. We can suppose that in any interesting case C gives the person strong support for belief in D, so  $P(D/C)$  is high, but if this term is not 1 then as it is multiplied by the terms that already exist in the  $P(C)$  expansion, it can only degrade the fidelity that  $P(C)$  had. The crucial new summands can add to the probability of D, and both have the term  $P(D/-C)$ , which is independent of all the other terms. When this term and the other such catch-all likelihoods that occurred in the calculation for the shorter proof of C are sufficiently high, then  $P(D) \geq P(C)$ .

A greater value for  $P(D/-C)$  means intuitively that a judgment that D is true has its level of justifiedness to some extent *whether or not* C is true. With Terell above, a relatively high value for this term corresponds to our being able to assume that if we are wrong that Terell is 90 % reliable he is still close to that. Here, we can interpret it as saying that in the inference to D one has somehow secured an extent of independence of justifiedness of a D-belief from a C-belief—even if C is not true D is sufficiently supported by the truth of B or by background assumptions. Another way of looking at it is that D is supported even if a mistake was made in inferring via the intermediate C.

In the extreme case of this, full independence of D from C, one would in making the inference from C to D have been essentially acquiring grounds that fully justified introducing a gap, replacing the inferences to and from C with nothing. The extreme case of strictly adding an inference that enhances legitimate confidence in the conclusion is the same as replacement of the last step of the original proof with a different step. What we are now discussing is essentially degrees of this replacement. Incidentally, this is the reason why the art buyer above who adds the step of consultation with his own expert is not necessarily increasing his potential error. To the extent that his degree of belief in the conclusion that the piece is real completely depends on his expert's opinion, that opinion screens off the previous process, rendering its steps probabilistically irrelevant. With that step he is

doing what I did by consulting Terrell above, but where I replaced one step (my judgment of my battery), the art buyer replaces a previous string of steps (other people's actions).

Realization of an extent of independence of one's result from the previous inference may also occur in situations which are not the next step of a lengthening of a proof, but intermediate steps in an existing proof. If one had undertaken a gap-filling process in order to secure a proof from A to D by smaller steps one could subsequently reverse that process, withdrawing the dependence of one's belief on those steps. One would replace the sequence of detailed steps by a secure leap from its beginning to its end if one discovered a degree of independence of that leap from these detailed steps.

This is a logical possibility which lines up nicely with a picture of how increasing justifiedness is sometimes secured. You begin with a hunch, or conviction, that D follows from A. Through filling in gaps (which is achieved by lengthening from A) you secure a path of justifiedness from A to D, secured in each step. Then you give up the reliance of your belief in D on those intermediate steps, and hold it only on the strength of belief in A. One might wonder, though, how the increased understanding that the gap-filling process gives could bring the required independence from those intermediate inferences? How could those steps that provided your justification in inferring D from A themselves justify your giving them up? Surely they alone cannot, but one possibility is where sinking into the details of the material, which would be necessary for filling in the gaps, enables one to see that there are multiple paths of reasoning from B to D, multiple ways of looking at the situation, multiple proofs, or multiple independent experiments. Then it would make sense that D is supported by B independently of whether one is right about C's being true or following from B.

I think this is one plausible picture of how arguments that are in substance enormously long and complicated can become more acceptably justified despite their vast number of supporting intermediate inferences. Another way of seeing how science and math are possible does not involve a change through which one realizes *after* the gap-filling process and new information that D's support by, say, A does not depend on the intervening steps. It comes from the values for the catch-alls, e.g.  $P(C/B)$ , being high already in the original proving because we come into the process of trying to verify our hunch already believing, on the basis of background knowledge, that C or B has some degree of plausibility.

One might wonder what the proving process could be contributing to our justification for believing D if we already were confident in the intermediate propositions when we started. But one thing that the proving did was to bring C and its probable support of D to our attention, showing that C, an already plausible proposition, is along the way. What we are now seeing is just that this newly discovered path needs independent support in its links in order to make its own support contribution. I think it should not surprise us that individual proofs or arguments cannot do everything for us. If we started with a hunch that D is supported by A and every line we came up with in forging a path to show it was a proposition that we otherwise and independently regarded as completely

implausible, I do not think we would feel this proof had provided a whole lot more security than we started with.

We just considered how reliability of our conclusion-belief can be enhanced by lengthening, gap-filling, or re-gapping via securing independence relations. What if we would be satisfied merely to preserve the reliability we had every time we added a step? What would the values of the new terms have to be to attain this? This will also answer the question whether the mechanism of enhancement just described or this preservation would be realistic; it will tell us how much independence from previous steps or how much independent plausibility for those previous steps themselves we need to secure. For ease of reference, I repeat the equations for the earlier arguments:

$$\Pr(C) = \Pr(C/B)\Pr(B/A)\Pr(A) + \Pr(C/-B)\Pr(-B/A)\Pr(A)$$

$$\Pr(D) = \Pr(D/C)\Pr(C/B)\Pr(B/A)\Pr(A) + \Pr(D/-C)\Pr(-C/B)\Pr(B/A)\Pr(A) \\ + \Pr(D/C)\Pr(C/-B)\Pr(-B/A)\Pr(A) + \Pr(D/-C)\Pr(-C/-B)\Pr(-B/A)\Pr(A)$$

As we said, in an argument worth considering  $P(D/C)$  will be high; one's new inference had a small potential error.  $\Pr(C/B)$ ,  $\Pr(B/A)$ ,  $\Pr(A)$  from the argument to C and re-occurring in that to D will also be high. Let us set all of these at .95. There are no terms with negations in their conditions except  $\Pr(C/-B)$  and  $\Pr(D/-C)$ , but these are enough to illustrate the point. They are independent of all other terms, and since the other inferences are now set, everything depends on these. I will consider two quantitative cases. In one we have  $\Pr(C/-B)$  at .5. C is 50 % likely (even) if B is not true. Then we have:

$$\Pr(C) = (.95)(.95)(.95) + (.50)(.05)(.95) = .88$$

What does  $\Pr(D/-C)$  have to be for the additional step to D to end with  $\Pr(D)$  at .88?

$$\Pr(D) = .88 = (.95)(.95)(.95)(.95) + \Pr(D/-C)(.05)(.95)(.95) \\ + (.95)(.50)(.05)(.95) + \Pr(D/-C)(.50)(.05)(.95) \\ = .82 + .02 + (.05 + .02)\Pr(D/-C)$$

$$\Pr(D/-C) = (.04)/(.07) = .57$$

If C is even halfway plausible independently of B, then D only needs to be somewhat more than halfway plausible in order to preserve the .88 reliability on adding a step to the proof. However, if C is not plausible independently of B, that is,  $\Pr(C/-B)$  is, say, .05, then,  $\Pr(D/-C)$  has to be .86 to make up for it. There is not space for further analysis here, but this calculation suggests that we need an average, independent, roughly halfway plausibility of each line in an individual argument in order to preserve reliability, or level of justified belief, over many inferences. This may be surprising because we do sometimes attach great significance to a single argument or proof. However, I do not think that significance needs to be denied; without forging the path an individual argument gives we might have nothing at all. It is only that in order for that argument to give what it can give we have to already have some relevant somewhat justified beliefs.



On the general question of how we can improve the reliability of our reasoning by adding steps, we have one more phenomenon to consider: double-checking an argument. The total probability analysis allows us to write down straightforwardly what I think is the intuitive explanation of the value of this practice. It depends on two comparisons of conditional probabilities, one set on the condition that there is no error in a given line, the other set on the condition that there is. For any kind of checking that we expect to reduce error, the probability given that there is no error, meaning that the inferred proposition is true, that you do not see an error the first time round is less than or equal to the probability you do not see an error both the first and second times round. That is, for proof-checking to succeed it must be that you do not have a tendency to see more errors where there are none the more times you look for them. This is because, or to the extent that, your method of checking reduces false negatives. However, on condition that an error exists, meaning that the inferred proposition is false, when we think checking will help we also presume that the probability that you do not see an error the first time around is lower than that you do not see an error both the first and second times round. Your checking method has some chance of catching false positives. The enhancement checking can bring is diminished if for a given error—positive or negative, you are as or more likely to make it the second time around as the first, say because you check it by a repetition of the very same procedure. But that is why we find it helpful to check inferences in the reverse order, and to ask others to check for us—these disrupt that repetitive quality of the checking.

All of these features can be brought out using total probability. To incorporate the contrast between what we get the first and second times around, consider an inference from A to B:

$$\Pr(B) = \Pr(B/A)\Pr(A)$$

Now consider adding in the conditions the information that we did not notice an error when we made the original argument, call this proposition  $N_1$ , and, separately, that we did not notice an error either in the original arguing or in the double-check, call this  $N_2$ :

$$\begin{aligned}\Pr(B/N_1) &= \Pr(B/A.N_1)\Pr(A/N_1) \\ \Pr(B/N_2) &= \Pr(B/A.N_2)\Pr(A/N_2)\end{aligned}$$

Focusing on the effect of double-checking the inference from A to B, we can understand there being an error as B being false while you become confident that it is true given that A is true (or the reverse, below). We can represent the greater probability of your finding an error given that there is an error, through double-checking than through single checking as:

$$\Pr(N_1/-B.A) > \Pr(N_2/-B.A).$$

In cases where we think double-checking would help, we also believe each type of checking is more likely to find an error if there is one than if there is not (more likely not to find an error,  $N_n$ , if there is not an error, B, than if there is, -B). That is:

$$\begin{aligned} \Pr(N_1/B.A) &> \Pr(N_1/-B.A) \text{ and} \\ \Pr(N_2/B.A) &> \Pr(N_2/-B.A) \end{aligned}$$

Finally, let us assume as noted above that if there is no error, B, then you are no more likely to (correctly) fail to report an error that is not there the first time around than you are after two times around:

$$\Pr(N_1/B.A) \leq \Pr(N_2/B.A)$$

Checking again does not make you more likely to imagine errors that are not there, but less. These four conditions correspond to the quality of any checking that can be expected to enhance the reliability of our conclusion.

We want to compare the probability of B, that is, that we make no error in asserting B, given that one check says no error, and give that two checks say no error. We expect the latter to be higher, and it is. For these four conditions and the assumption that  $\Pr(B) > 0$  combine to imply that  $\Pr(-B/N_2) < \Pr(-B/N_1)$ ,<sup>9</sup> that is, that  $\Pr(B/N_2) > \Pr(B/N_1)$ . B is more likely given that two checks said there was no error than given that only one check said no error. It also follows that the greater the ratio of  $\Pr(N_1/-B)/\Pr(N_1/B)$  to  $\Pr(N_2/-B)/\Pr(N_2/B)$ , the greater is  $\Pr(B/N_2)$  than  $\Pr(B/N_1)$ . This means that the *more* sensitive that doing two checks makes you to error than the first check did, the less potential error in your conclusion, B.

## 5 Conclusion

It is commonly believed that potential error necessarily builds up over the course of inferences and an increase of premises in a proof, because in adding more steps we are adding more sources of error and error is presumed to grow by the conjunction rule. This is sometimes greeted with dismay, as it should be because we rely, heavily if only implicitly, on complicated chains of dependency of our beliefs on other beliefs. The appearance that potential error builds up over steps of inference would also have the absurd implication that consulting expert sources, double-checking arguments, adding of evidence for a conclusion, and filling gaps in arguments necessarily reduce our fidelity.

The claim of this paper has been that the common belief that error necessarily grows with steps is false. It is thought true because of a mistaken use of the conjunction rule to determine the reliability of a conclusion-belief that is not a conjunction. When we calculate the rationally justified degree of belief in the conclusion of an argument probabilistically and correctly, by the total probability rule, we see multiple circumstances under which addition of steps preserves or enhances reliability and why. Multiple arguments, prior plausibility of claims in the proof, and replacement are all realistic mechanisms for the possible error to be lower despite a greater number of steps, and they explain many intuitive features of

<sup>9</sup> This can be derived beginning with total probability on  $N_1$  and on  $N_2$  (via B)—which yields that if  $\Pr(B) > 0$  then  $\Pr(N_1)/\Pr(N_2) < \Pr(N_1/-B)/\Pr(N_2/-B)$ . This combined with Bayes' theorem applied to  $\Pr(-B/N_1)$  and  $\Pr(-B/N_2)$  gives that  $\Pr(-B/N_1) > \Pr(-B/N_2)$ .

the phenomenon of building longer and longer arguments, and extra consultations, to increase our knowledge.

**Acknowledgments** This study was supported in large part by NSF grant SES-0823418.

## References

- Detlefsen, Michael, & Luker, M. (1980). The four-color theorem and mathematical proof. *The Journal of Philosophy*, *LXXVII*, 803–820.
- Garber, D. (1983). Old evidence and logical omniscience in bayesian confirmation theory. In J. Earman (Ed.), *Testing scientific theories* (Vol. X, pp. 99–131). Minneapolis: Minnesota Studies in the Philosophy of Science.
- Goldman, A. (2008). What is Justified Belief? In *Epistemology: An anthology*, 2nd edn. (pp. 333–347. Malden: Blackwell Publishing.
- Hacking, I. (1967). Slightly more realistic personal probability. *Philosophy of Science*, *34*(4), 311–325.
- Jeshion, R. (1998). Proof checking and knowledge by intellection. *Philosophical Studies*, *92*, 85–112.
- Kitcher, P. (1984). *The nature of mathematical knowledge*. Oxford: Oxford University Press.
- Resnik, M. D. (1989). Computation and mathematical empiricism. *Philosophical Topics*, *XVII*, 129–143.
- Vogel, J. (2006). Externalism resisted. *Philosophical Studies*, *131*, 729–742.