

# The Bayesian explanation of transmission failure

Geoff Pynn

Received: 1 October 2010 / Accepted: 7 February 2011  
© Springer Science+Business Media B.V. 2011

**Abstract** Even if our justified beliefs are closed under known entailment, there may still be instances of transmission failure. Transmission failure occurs when  $P$  entails  $Q$ , but a subject cannot acquire a justified belief that  $Q$  by deducing it from  $P$ . Paradigm cases of transmission failure involve inferences from mundane beliefs (e.g., that the wall in front of you is red) to the denials of skeptical hypotheses relative to those beliefs (e.g., that the wall in front of you is not white and lit by red lights). According to the Bayesian explanation, transmission failure occurs when (i) the subject's belief that  $P$  is based on  $E$ , and (ii)  $P(Q|E) < P(Q)$ . But there are compelling cases of transmission failure where  $P(Q|E) > P(Q)$ . No modifications of the Bayesian explanation are capable of accommodating such cases, so the explanation must be rejected as inadequate. Alternative explanations employing simple subjunctive conditionals are fully capable of capturing all of the paradigm cases, as well as those missed by the Bayesian explanation.

**Keywords** Transmission · Closure · Justification · Bayesian epistemology

## 1 Transmission failure

You can often enlarge your set of justified beliefs by deduction alone. Glancing at the table, I see my wife's keys. I justifiably believe that her keys are on the table, and then come to believe (1) by a straightforward deductive inference:

(1) She doesn't have the keys in her purse.

---

G. Pynn (✉)  
Department of Philosophy, Northern Illinois University, 915 Zulauf Hall, DeKalb, IL 60115, USA  
e-mail: [gpynn@niu.edu](mailto:gpynn@niu.edu)

Thus do I acquire a justified belief in (1). But while (2) also follows from my belief that my wife's keys are on the table, it doesn't seem that I could come to justifiably believe it in the same way:

(2) The keys on the table aren't fakes left there by my wife to mislead me.

My justification for believing that my wife's keys are on the table "fails to transmit" to (2).<sup>1</sup>

If transmission fails from  $P$  to  $Q$ , deducing  $Q$  from  $P$  doesn't enable you to acquire a justified belief in  $Q$ . The existence of cases of transmission failure does not imply that your set of justified beliefs isn't closed under known entailment. Suppose that they are; i.e., that whenever you have a justified belief in  $P$  and know that  $P$  entails  $Q$ , you also have a justified belief in  $Q$ . This supposition is compatible with there being some pairs of justified beliefs  $P$  and  $Q$  where you know that  $P$  entails  $Q$ , but where you could not have acquired your justified belief in  $Q$  by deduction from  $P$ . Transmission failure is compatible with a range of closure principles.<sup>2</sup>

Among those who accept that transmission failure is a genuine phenomenon, opinions vary as to what explains it. In a series of papers, Crispin Wright has developed two influential accounts of transmission failure.<sup>3</sup> More recently, epistemologists have suggested that we can instead understand transmission failure in probabilistic terms. According to the "Bayesian explanation" for transmission failure, when  $E$  is your evidence for  $P$ , you cannot acquire a justified belief in  $Q$  by deduction from  $P$  when  $P(Q|E) < P(Q)$ ; i.e., when  $E$  lowers the probability of  $Q$ .

My purpose here is to examine the Bayesian explanation. After looking at some paradigm cases of transmission failure (Sect. 2), I'll lay out the Bayesian explanation

<sup>1</sup> The term 'transmission failure' is from Crispin Wright (see Wright 1985, 2000, 2002, 2003, 2008). Recent discussions include Davies (2003, 2004), Neta (2007), Okasha (2004), Pryor (2004, 2009), Silins (2005, 2007), Smith (2009).

<sup>2</sup> Wright (2000) makes this point as well. I should note that the distinction between "transmission" and "closure" is murky, and that my characterization of transmission failure renders the phenomenon incompatible with *some* "closure" principles. Consider, for example, this "closure" principle from Hawthorne:

If one knows  $P$  and competently deduces  $Q$  from  $P$ , thereby coming to believe  $Q$ , while retaining one's knowledge that  $P$ , one comes to know that  $Q$  (Hawthorne 2004, p. 29).

If a principle like Hawthorne's is correct, then it is plausible that one can always gain a justified belief in  $Q$  by deduction from  $P$  (provided the other conditions in the antecedent of the principle are satisfied), and hence it is unlikely that there are any genuine cases of transmission failure as I have characterized it (see Sect. 5 of Silins 2005 for a persuasive argument to this end). To save a principle like Hawthorne's while allowing genuine transmission failure, one could give a weaker definition of transmission failure than I have. One might say that transmission fails from  $P$  to  $Q$  when the subject's justification for believing  $P$  cannot contribute to her justification for believing  $Q$ , though she can acquire a justified belief in  $Q$  by deduction from  $P$ . I have no strong objection to such a characterization, though I do find the suggestion that a belief could be justified by a factor which is purely independent of its basis to be somewhat troubling. In any event, none of the arguments in the present paper turn on the resolution of these issues or on the precise characterization of transmission failure.

<sup>3</sup> See especially Wright (2000, 2002, 2003). Brown (2004) suggests repairs to Wright's account. Some authors have attempted to give a probabilistic version of Wright's account (see Okasha 2004 and Chandler 2010), as well as unpublished work by Luca Moretti. Okasha attempts to give a Bayesian analysis of Wright's distinction between "direct" and "indirect" support. See note 12 below for brief discussion of Chandler's account.

(Sect. 3). Then I'll describe a case of transmission failure that the Bayesian explanation can't handle (Sect. 4). Analogous cases are simple to construct; the Bayesian explanation misses a large swath of instances of the phenomenon it is supposed to explain. After considering some possible responses on behalf of the Bayesian (Sect. 5), I'll say where I think it goes wrong (Sect. 6).

## 2 Paradigm cases

Our best evidence for the existence of transmission failure comes from cases. There are many famous triads where, supposing that one's evidence for  $P$  is  $E$ , it seems one cannot gain justification for believing one of  $P$ 's entailments  $Q$  by deduction from  $P$ . Here is Dretske's famous example:

RED

$E_R$  That wall appears to be red.

$P_R$  That wall is red.

$Q_R$  That wall is not white and illuminated by red lights.<sup>4</sup>

If  $E_R$  is the basis for your justified belief in  $P_R$ , then merely deducing  $Q_R$  from  $P_R$  seems clearly not to be a way of acquiring a justified belief in  $Q_R$ .<sup>5</sup>

Here's another famous case from Dretske (1970):

ZEBRA

$E_Z$  Those animals are horse-shaped, striped, in a pen marked 'Zebras', etc.

$P_Z$  Those are zebras.

$Q_Z$  Those aren't mules cleverly disguised by the zoo authorities to look like zebras.

And the ubiquitous:

BIV

$E_B$  It looks and feels like I have hands.

$P_B$  I have hands.

$Q_B$  I'm not a bodiless brain in a vat being fed hand-ish experiences.

In each triad, the first proposition is defeasible evidence sufficient to justify belief in the second, and the second entails the third. But it is extremely implausible to say that someone's belief in the third could be justified by deduction from the second. These are paradigm examples of transmission failure. Any account of the phenomenon should either predict that transmission fails in these cases, or else explain why it so appears.

Note that the status of each case as an example of transmission failure depends essentially upon  $E$ . Suppose that  $P_R$  was justified not by  $E_R$  but by your memory of having painted the wall red this morning. Or that  $P_Z$  was justified by your knowledge of recent DNA tests on the animals before you. In either case, the appearance of

<sup>4</sup> This example is originally due to Dretske (1970) and has been discussed more recently by Cohen (2002), Pryor (2004), Wright (2002, 2008) and others.

<sup>5</sup> Pryor comes close to denying this. He writes that your  $E_R$ -based justification to believe  $P_R$  "contributes to the credibility of the claim that the wall isn't white and lit by tricky lights" (Pryor 2004, p. 362). Still, Pryor recognizes that *something* is fishy with the argument represented by Red, and attempts to provide an alternative explanation for what's gone wrong.

transmission failure would disappear. Transmission failure does not occur from  $P$  to  $Q$  *simpliciter*, but only relative to one's evidence for  $P$ .

These cases needn't be seen as counterexamples to closure. Suppose that if you have a justified belief in  $P_R$ , you must also have a justified belief in  $Q_R$ . If transmission fails in Red, then your justification for  $Q_R$  must come from some source other than deduction from  $P_R$ . But there are many candidate sources. Perhaps you are justified in believing that there are no red light bulbs in the house you are in. Or that white walls convincingly lit by red lights are rare. Or maybe your justification comes from some default entitlement to accept the proposition that things are as they appear.

Still, retaining closure while acknowledging transmission failure does invite skeptical danger. Consider BIV. If justified beliefs are closed under known entailment, then you have a justified belief that you have hands only if you have a justified belief that you are not a bodiless brain in a vat. But if transmission fails from your perceptually-justified belief that you have hands to  $Q_B$ , surely it fails from any such perceptually-justified belief. So your justification for  $Q_B$  can't come from any perceptually-justified beliefs that entail it. This makes it difficult to see where it *could* come from, and hence how your belief that  $Q_B$  could be justified.<sup>6</sup>

My concern here is not to address this skeptical worry, but to understand what goes wrong in cases like BIV, Zebra, and Red. Why is it that beliefs in the various  $Q$ 's can't be justified by deduction from the relevant  $P$ 's? Or, at least, why does it seem that this is the case?

### 3 The Bayesian explanation

According to the Bayesian explanation, when your justification for  $P$  comes from  $E$ , transmission fails from  $P$  to  $Q$  when  $Pr(Q|E) < Pr(Q)$ .<sup>7</sup>

Given very plausible assumptions, we can show that the explanation correctly predicts that each of the paradigm cases is an example of transmission failure.

If the wall *is* white and illuminated by red lights, then it appears to be red; i.e.,  $\neg Q_R$  entails  $E_R$ .<sup>8</sup> So,

$$Pr(E_R|\neg Q_R) \approx 1$$

<sup>6</sup> One response is to deny that BIV is a case of transmission failure. This is how I understand Pryor's position. Another is to say that one's justification for  $Q_B$  comes from a non-perceptual source.

<sup>7</sup> Silins (2007) offers the Bayesian explanation for why transmission fails in BIV. Hawthorne (2004, fn. 42), (Cohen, 2005, pp. 424–425), and White (2006) make similar suggestions. The only  $E$ – $P$ – $Q$  triads that are candidates for transmission failure are cases where  $E$  is typically sufficient to give you justification to believe  $P$ . Clearly, then, there will be other necessary conditions in addition to  $Pr(Q|E) < Pr(Q)$ ; e.g., that  $Pr(P|E) > Pr(P)$ , or that  $Pr(P|E)$  meet or exceed some threshold of rational acceptability.

<sup>8</sup> Properly speaking, of course,  $\neg Q_R$  does not entail  $E_R$ ; if there is a film of yellow plastic wrap covering the wall, it will look orange, not red, when lit by red lights. This loose use of 'entails' is common in the literature, and is required for the Bayesian explanation to get off the ground. Troubled readers can preface 'entails' with 'together with relevant background assumptions, ...'.

In addition,  $E_R$  was not certain in advance. The wall might have appeared green instead of red. Thus,

$$Pr(E_R) < 1$$

Hence  $Pr(E_R|\neg Q_R) > Pr(E_R)$ . By Bayes' Theorem,  $Pr(\neg Q_R|E_R) > Pr(\neg Q_R)$ .<sup>9</sup> So,

$$Pr(Q_R|E_R) < Pr(Q_R)$$

Similar reasoning will show that analogous inequalities are satisfied by Zebra and BIV.

Why should this inequality lead to transmission failure? We understand  $Pr$  as a function from propositions to a subject's credences, and assume that when  $Pr(P|E) < Pr(P)$ , the subject ought rationally to lower her confidence in  $P$  on learning  $E$ . Thus in Red she ought rationally to lower her confidence in  $Q_R$  on learning  $E_R$ . But how could she gain justification to believe  $Q_R$  by acquiring evidence that rationally enjoins her to lower her confidence in  $Q_R$ ? Learning  $E_R$  couldn't directly justify belief in  $Q_R$  if the rational response to learning  $E_R$  is to become less confident in  $Q_R$ . Could learning  $E_R$  gave her indirect justification for  $Q_R$  by giving her justification for  $P$ ? If it did, then it would still be the case that  $E_R$  would justify belief in  $Q_R$  despite rationally enjoining her to become less confident in  $Q_R$ . And this seems implausible. So goes the Bayesian explanation.<sup>10</sup>

Is it correct that any evidence which rationally enjoins you to lower your confidence in  $P$  couldn't also justify you in believing  $P$ ? Maybe not always. I am holding a lottery ticket and on the basis of an unjustified hunch I am certain that the odds of winning are one in a hundred million. Letting  $P = \text{that my ticket will lose}$ ,  $Pr(P) = 0.99999999$ , given my belief about the odds. Since I have no evidence beyond my hunch that supports  $P$ , my belief that  $P$  is unjustified. Now I learn from an eminently trustworthy source that the odds of winning are, in fact, one in ten million. Call this proposition  $E$ . By learning  $E$ , I gain justification to believe  $P$ ; my belief that  $P$  is now justified. But letting  $E = \text{that the odds of winning are one in ten million}$ ,  $Pr(P|E) < Pr(P)$ ; I ought to lower my confidence in  $P$  upon learning  $E$ .

Still, the case of the lottery hunch is unusual in that one's degree of confidence in  $P$  prior to learning  $E$  is unjustified. Our paradigm cases of transmission failure do not involve a transition from a previously unjustified degree of confidence in  $Q$ .

Here is a second potential difficulty. Suppose that when I have a visual experience with the content  $P$ , I thereby learn  $P$  (at least when various background conditions are satisfied), and that when I learn  $P$ , I am justified in setting  $Pr(P)$  to 1. And suppose

<sup>9</sup> I assume throughout that  $Pr(\neg Q_R) > 0$ . The assumption is apt: it was not certain in advance that the wall would not be white and illuminated by red lights.

<sup>10</sup> Cf. Silins (2007). Silins argues that Moore's (fictional) inference from his experientially-justified belief that he has hands to the conclusion that he is not a bodiless brain in a vat having hand-like experiences is an instance of transmission failure, since, as he puts it: "[I]f Moore's experience makes his justification to believe [ $Q_R$ ] go up, then it can't be that his confidence in [ $Q_R$ ] should instead go down." Contraposing: if his confidence in  $Q_R$  should go down, his experience does not make his justification to believe  $Q_R$  go up.

that when the wall appears to be red I have a visual experience with the content *that wall is red*. Since  $P_R$  entails  $Q_R$ ,  $Pr(Q_R|P_R) = 1$ . Thus when the wall looks red to me, I learn that the wall is red, am thereby justified in setting  $Pr(P_R)$  to 1, and ought rationally to set  $Pr(Q_R)$  to 1, as well. If this is the correct way to understand the epistemology of visual experience, then it seems that the rational response to the wall's looking red to me is not for my confidence in  $Q_R$  to go down, but rather to go up. But, one might think, so much the worse for this picture of the epistemology of visual experience. And in any event, it is easy to construct cases of transmission failure where  $E$  is not plausibly regarded as the content of a visual experience.

A more serious problem is raised by Pryor (2009). Consider the following variant of BIV:

$E_B$  It looks and feels like I have hands.

$P_B$  I have hands.

$R_B$  I have hands and I'm not a bodiless brain in a vat being fed handish experiences.

The intuition of transmission failure from  $P_B$  to  $R_B$  is nearly as strong as that in the original case from  $P_B$  to  $Q_B$ . But  $R_B$  is logically equivalent to  $P_B$ ; thus  $Pr(P_B|E) = Pr(R_B|E)$ . So the Bayesian explanation fails to capture this variant. Now, one might reply by pointing out that  $R_B$  is nothing more than the conjunction of  $P_B$  and the original  $Q_B$ , and that the intuition of transmission failure from  $P_B$  to  $(P_B \& Q_B)$ , though a misfire, is harmless enough given that there is genuine transmission failure from  $P_B$  to  $Q_B$ . But this case portends bigger trouble for the Bayesian. In the next section, I shall consider some more general difficulties for her explanation along these lines.

## 4 Problem cases

The Bayesian explanation fulfills one key desideratum of an explanation for transmission failure: it accurately predicts the paradigm cases discussed in Sect. 2. But there are other intuitive cases of transmission failure it does not predict. A minor tweak to the account can capture the first family of cases, but there is no obvious fix for the second.

### 4.1 Transmission failure when $Pr(Q|E) \approx Pr(Q)$

When  $E$  and  $\neg Q$  are independent,  $Pr(\neg Q|E) \approx Pr(\neg Q)$ , so  $Pr(Q|E) \approx Pr(Q)$ , and hence it is not the case that  $Pr(Q|E) < Pr(Q)$ . But there are intuitive cases of transmission failure where  $E$  and  $\neg Q$  are independent. For example:

MEG

$E_M$  Meg's train was scheduled to arrive at six o'clock, the station is a few blocks from here, and she's planning to walk over.

$P_M$  Meg will be here shortly.

$Q_M$  Meg wasn't just rushed to the hospital after collapsing of a heart attack.

If your belief in  $P_M$  is justified by  $E_M$ , you can't acquire a justified belief in  $Q_M$  by deducing it from  $P_M$ . But  $Q_M$  is unrelated to  $E_M$ : unless Meg's heart is unusually

weak, or some other extraordinary background condition obtains, the probability of  $Q_M$  is unaffected by  $E_M$ . So  $Pr(Q_M|E_M) \approx Pr(Q_M)$ . Thus the Bayesian explanation does not predict that Meg is a case of transmission failure.

It is very easy to construct cases like Meg. Begin with some evidence  $E$  that does not entail  $P$  but which would ordinarily justify belief in  $P$ . Then imagine a scenario  $S$  such that (i)  $S$  entails  $\neg P$ , (ii)  $S$  and  $E$  are independent, and (iii) if  $S$  were true, you could easily falsely believe  $P$  on the basis of  $E$ . Let  $Q = \neg S$ . The resulting  $E$ - $P$ - $Q$  triad will elicit a strong intuition of transmission failure, but it will not be the case that  $Pr(Q|E) < Pr(Q)$ .

However, a modest tweak to the Bayesian explanation enables it to capture cases like this. We simply replace the “ $<$ ” with “ $\lesssim$ ”: an  $E$ - $P$ - $Q$  triad is a case of transmission failure only if  $Pr(Q|E) \lesssim Pr(Q)$ . Meg satisfies this condition, as will any case where  $E$  and  $Q$  are independent. But a more serious problem is at hand.

#### 4.2 Transmission failure when $Pr(Q|E) > Pr(Q)$

We can also construct cases of transmission failure where  $E$  raises the probability of  $Q$ .

##### COLORED

You are in a house with 100 rooms. 49 are painted white, and 51 are painted red. Of the white rooms, one is lit by red lights, and four others are lit by lights of other colors. All of the red rooms are lit by white lights. You’re about to open the door to the first room in the house.

Opening the door to the first room, you notice that it looks red and reason as follows:

$E_R$  That wall appears to be red.

$P_R$  That wall is red.

$R_R$  That wall is not white and illuminated by colored lights.

Colored appears to be a case of transmission failure. If your belief in  $P_R$  is justified by  $E_R$ , you cannot acquire a justified belief in  $R_R$  by deduction from  $P_R$ . But the Bayesian explanation fails to predict as much. Since 50 of the 100 rooms look red,

$$Pr(E_R) = 0.5$$

Of the five white rooms lit by red or colored lights, only one looks red. So,

$$Pr(E_R|\neg R_R) = 0.2$$

Thus  $Pr(E_R|\neg R_R) < Pr(E_R)$ . By Bayes’ Theorem,  $Pr(\neg R_R|E_R) < Pr(\neg R_R)$ , and hence  $Pr(R_R|E_R) > Pr(R_R)$ . The wall’s appearing red raises the probability that it’s not white and lit by colored lights.

Colored is by no means an isolated example. Many  $E$ - $P$ - $Q$  triads that exhibit transmission failure where  $Pr(Q|E) < Pr(Q)$  can be turned into such cases. First, choose some  $R$  such that (i)  $P$  entails  $R$ , (ii)  $R$  entails  $Q$ , but not vice versa, and (iii)  $R$  makes  $Q$  psychologically salient. For example, you can replace Zebra’s  $Q_Z$  with:

$R_Z$  Those are neither mules disguised like zebras nor deer disguised like gazelles. Then specify the relevant background information so that  $Pr(E|\neg R)$  is relatively low. In this case, we might suppose that the zoo frequently disguises deer as gazelles, so that  $\neg R_Z$  makes it likely that the animals in front of you look like gazelles, not zebras. *Et voilà!* An intuitive case of transmission failure which the Bayesian explanation is unable to predict.

## 5 Bayesian responses

The proponent of the Bayesian explanation could respond to a case like Colored in one of two ways. The first is to deny that it is a case of transmission failure. The second is to attempt to accommodate it with a more drastic modification of the explanation. I'll consider each response in turn.

### 5.1 Is Colored a case of transmission failure?

Could it be that Colored is not a case of transmission failure? If it's not, then the fact that the Bayesian explanation fails to predict as much is a virtue, not a cost. But this is a difficult position to maintain: it's hard to deny that Colored is a case of transmission failure while also maintaining that Red *is* one. And clearly it's not open to the proponent of the Bayesian explanation to deny that Red is a case of transmission failure.

Suppose that transmission fails in Red but not in Colored. Now consider the following argument:

$E_R$  That wall appears to be red.

$P_R$  That wall is red.

$R_R$  That wall is not white and illuminated by colored lights.

$Q_R$  That wall is not white and illuminated by red lights.

Does this argument specify a route to acquiring a justified belief in  $Q_R$ ? If the answer is "no," then the step at which transmission fails must be from  $R_R$  to  $Q_R$ , since (given our supposition) it does not fail before then. But this is quite implausible. "I can deduce from its being red that it isn't white and lit by colored lights. But I can't go on to conclude that it's not white and lit by red lights!"

On the other hand, if the answer is "yes," then while you can't acquire a justified belief in  $Q_R$  by deduction from  $P_R$ , you *can* acquire one by a more circuitous deduction that detours through the (stronger!)  $R_R$ . This, too, is quite implausible.

### 5.2 Modification I: infection

The alternative to denying that Colored is a case of transmission failure is to modify the Bayesian explanation in a way that captures such cases.

Note that  $Q_R$  is entailed by  $R_R$ . Perhaps the transmission failure exhibited by Red somehow "infects" Colored. How could we modify the Bayesian explanation to capture this infection? Here is an obvious suggestion. Let us suppose that an  $E-P-R$



triad is an instance of transmission failure just in case either (i)  $Pr(R|E) \lesssim Pr(R)$  (ordinary cases) or (ii)  $R$  entails some  $Q$  such that  $Pr(Q|E) \lesssim Pr(Q)$  (infection cases). This account would accurately predict that Colored is a case of transmission failure.

But the proposal is far too strong. If we allow it, then virtually any  $E$ - $P$ - $R$  triad will count as an infection case. For example:

- ORANGE
- $E_O$  That wall appears to be orange.
- $P_O$  That wall is orange.
- $R_O$  That wall is not yellow.

Orange is clearly not a case of transmission failure; deduction from  $P_O$  is a perfectly good way to acquire a justified belief in  $R_O$ . However,  $R_O$  entails  $Q_O$ :

$Q_O$  That wall is not yellow and illuminated by red lights.

And  $Pr(Q_O|E_O) < Pr(Q_O)$ , for reasons that are by now familiar. Thus the proposed modification incorrectly implies that Orange is a case of transmission failure.

### 5.3 Modification II: dragging

In his unpublished manuscript “Dragging and Confirming,” Matthew Kotzen argues that when  $E$  confirms  $P$ ,  $E$  also confirms  $P$ ’s entailment  $Q$  only if  $Pr(Q) < Pr(P|E)$ . Kotzen calls this the dragging condition. He suggests that in cases of transmission failure, the dragging condition is not satisfied. Note that if  $P$  entails  $Q$  and  $Pr(Q|E) < Pr(Q)$ , the dragging condition *can’t* be satisfied, so any case predicted by the Bayesian explanation to be a case of transmission failure will also fail to satisfy the dragging condition.<sup>11</sup> But a triad could fail to satisfy the dragging condition even if  $Pr(Q|E) > Pr(Q)$ . Perhaps, then, an account that incorporates the dragging condition will capture what is wrong in Colored without being overly strong. Following Kotzen, let us propose that an  $E$ - $P$ - $Q$  triad is a case of transmission failure when the dragging condition is not satisfied; i.e., when  $Pr(P|E) \leq Pr(Q)$ .

Unfortunately, the proposal does not work with Colored. You know that 95 of 100 rooms are not white and lit by colored lights, so

$$Pr(R_R) = 0.95$$

The chances of a red-looking room’s being red are 49 in 50, so

$$Pr(P_R|E_R) = 0.98$$

Thus  $Pr(R_R) < Pr(P_R|E_R)$ ; the dragging condition is satisfied. So the proposal is too weak.<sup>12</sup>

<sup>11</sup> Since  $P$  entails  $Q$ ,  $Pr(P|E) \leq Pr(Q|E)$ . Thus, if  $Pr(Q|E) < Pr(Q)$ ,  $Pr(P|E) < Pr(Q)$ .

<sup>12</sup> This is no complaint against the dragging condition itself, which Kotzen’s fine paper has convinced me is necessary for  $E$  to confirm  $P$ ’s entailments; i.e., for  $E$  to function as evidence for  $P$ ’s entailments.

## 6 Where things stand

If the Bayesian explanation is thought of as an attempt to state a true sufficient condition for transmission failure, then it is unscathed by the cases discussed above. None shows that transmission can succeed even if  $Pr(Q|E) > Pr(Q)$ . But if the Bayesian explanation is seen as an attempt to state a true necessary condition for transmission failure, then Colored undermines its adequacy. Given the failure of the two “patches” to the explanation that I considered above, it is not clear how the Bayesian explanation could be modified to restore its adequacy.

The non-necessity of  $Pr(Q|E) < Pr(Q)$  to cases of transmission failure undermines the claim that the presence of the inequality *explains* why transmission fails. Giving a sufficient condition for a fact is not explaining that fact. That the curfew is later than 10PM is sufficient for its being later than 9PM, but it does not explain why it is later than 9PM. One condition explains another only when the second is the case because of the first; if the first is unnecessary for the second, it is harder to make the case that the second is the case because of the first. The existence of transmission failure when  $Pr(Q|E) > Pr(Q)$  weakens the claim that transmission fails in our paradigm cases because  $Pr(Q|E) < Pr(Q)$ . Even if the presence of the inequality is predictive of transmission failure, that the inequality is unnecessary suggests that it is not explanatory.

It is open to the proponent of the Bayesian explanation to say that while the fact that  $Pr(Q|E) < Pr(Q)$  does not explain transmission failure in Colored, the presence of the inequality does explain why transmission fails in the paradigm cases. It is possible for different instances of a one phenomenon to have distinct explanations. This response would be satisfying only if no unified explanation for the phenomena were available. In the next section I will suggest two such alternatives.

## 7 Two explanations involving subjunctive conditionals

The alternatives both feature a subjunctive conditional. Suppose that transmission of justification from an  $E$ -based belief that  $P$  to  $Q$  is blocked when the following conditional is false:

$S$  If  $Q$  were false,  $E$  would be false.<sup>13</sup>

In each of our paradigm cases, if  $Q$  were false,  $E$  would be true, and so  $S$  is false. In Red, the wall would appear to be red if it were white and illuminated by red lights.

---

Footnote 12 continued

The lesson of Colored is that  $E$  can confirm  $P$ 's entailment  $Q$  even if you can't acquire a justified belief in  $Q$  by deduction from  $P$ .

Note, too, that Colored is a counterexample to Chandler (2010)'s proposal that in cases of transmission failure it must be that  $Pr(P|E) > t \supset Pr(Q) > t$ , where  $t$  marks a “threshold sufficient for rational acceptability.” Let  $t = 0.97$ . Then  $Pr(P_R|E_R) > t$ , but  $Pr(R_R) < t$ . If this is too low a threshold, raise the total number of rooms in the example. Provided  $t < 1$ , the details of Colored can be specified so that Chandler's conditional remains unsatisfied.

<sup>13</sup> Some readers may find it easier to interpret the conditional if stated as:  $Q$  would not be true unless  $E$  were true.

In Zebra, those animals would be horse-shaped, striped, and so on if they were mules cleverly disguised to look like zebras. And in BIV, it would look and feel like I have hands if I were a bodiless brain in a vat. Thus in each of our paradigm cases, *S* is false.

In Meg, we have the same result. If Meg were just rushed to the hospital after collapsing of a heart attack, her train would still have been scheduled to arrive at six o'clock, the station would still have been a few blocks away, and so on. So here, too, *S* is false. In Colored, matters are somewhat more delicate. Given the background of the case, it does not seem true that if the wall were white and lit by colored lights, it would appear red. If the wall were white and lit by colored lights, it might look to be some other color than red. But it is clear enough, given the background of the case, that if the wall were white and lit by colored lights, it *might* look red. So if the following argument is valid, Colored satisfies the condition under consideration:

- (3) If the wall were white and lit by colored lights, it might look red.  
 ∴ It is false that: if the wall were white and lit by colored lights, it would not look red.

Is (3) valid? It is commonly assumed that “might”-conditionals are the duals of the corresponding “would”-conditionals; i.e., that “If *A* were true, *C* would be true” is equivalent to “It is not the case that if *A* were true, *C* might not be true”.<sup>14</sup> The “duality thesis,” as this assumption is often known, implies that (3) is valid. Given the duality thesis, the present condition predicts that Colored is a case of transmission failure.<sup>15</sup>

But the duality thesis is controversial. Another view has it that a “might”-conditional expresses the possibility that the corresponding “would”-conditional is true.<sup>16</sup> On this reading, (3) is not valid. Instead, (4) is valid:

- (4) If the wall were white and lit by colored lights, it might look red.  
 ∴ It is possible that: if the wall were white and lit by colored lights, it would look red.

What sort of possibility is at issue here? The most common idea is that it is epistemic possibility.<sup>17</sup> On this understanding of the “might” conditional, in Colored it is epistemically possible that the wall would look red if it were white and lit by colored lights. This suggests an explanation of transmission failure involving *S*'s contradictory. Suppose that an *E–P–Q* triad is an instance of transmission failure whenever *S*\* is epistemically possible:

<sup>14</sup> Lewis's semantics for counterfactual conditionals assumes the duality thesis (see Lewis 1973, p. 21). While the antecedent of *S* contains a “were”, I am not suggesting that the conditional should be read as a counterfactual.

<sup>15</sup> If we read *S* as a counterfactual and interpret it using the possible worlds semantics in Lewis (1973), then the proposal here is closely related to the possible worlds account of transmission failure for knowledge given by Smith (2009). However, this interpretation invites the standard sorts of objections raised against counterfactual analyses of epistemic phenomena; to wit, counterexamplifying by cases where subjects in “nearby” possible world are in a different epistemic situations than that of the actual subject.

<sup>16</sup> For discussions of this rival view see Stalnaker (1984) and DeRose (1994). Unpublished work by Matt Benton was helpful to me here.

<sup>17</sup> The version of the view DeRose discusses reads the “might” as expressing epistemic possibility. Stalnaker's view is less straightforward; for him the “might” can express different sorts of possibility in different contexts, though normally it is epistemic.

$S^*$  If  $Q$  were false,  $E$  would be true.

This explanation correctly predicts that Colored is a case of transmission failure, and it fares just as well with our paradigm cases, as well. It is surely epistemically possible that if the wall were white and lit by red lights, it would look red; that those animals would be horse-shaped, striped, and so on if they were mules cleverly disguised to look like zebras; and that it would look and feel like I have hands if I were a bodiless brain in a vat.

So now we have two alternatives to the Bayesian explanation, both of which capture the paradigm cases and Colored with a single condition. The first assumes the duality thesis and says that transmission fails when  $S$  is false. The second is independent of the duality thesis and says that transmission fails when  $S^*$  is epistemically possible. Obviously, each stands in need of refinement, further testing, and supplementation with a deeper account of *why* transmission fails when the proposed condition is satisfied.<sup>18</sup> Even without that work, the apparent availability of simple, superior alternatives to the Bayesian explanation completes the present case against it.

**Acknowledgments** Thanks to audiences at the 2009 Central States Philosophical Association Meeting and the 2010 Central APA Division Meeting for good questions, and in particular to Paul Weirich and Kelly Becker, who provided charitable and helpful comments on those occasions. Thanks also to David Manley for a helpful conversation, Matt Kotzen for permission to cite his manuscript, Matt Benton for help with “would” and “might” conditionals, and Chris Tucker for generously reading a complete draft and engaging in fruitful correspondence. Finally, thanks to two anonymous referees for uncommonly insightful, sympathetic, and productive suggestions.

<sup>18</sup> An anonymous reviewer makes the following interesting suggestion. Let’s characterize an *undermining defeater* relative to  $E$  and  $P$  as any evidence against the claim that if  $P$  were false,  $E$  would be false (this roughly follows (Pollock 1974, pp. 42–43). The suggestion is that when  $S$  is true or  $S^*$  is epistemically possible, given that  $P$  entails  $Q$ ,  $\neg Q$  is an undermining defeater relative to  $E$  and  $P$ . Then my proposal can be recast as follows: justification from an  $E$ -based belief that  $P$  fails to transmit to  $Q$  when  $\neg Q$  is an undermining defeater relative to  $E$  and  $P$ .

I hesitate to endorse this suggestion for two reasons. The first is that Pryor argues nicely against a very similar approach, which he calls the “Anti-Underminer Model” of transmission failure, in the last section of Pryor (2009). His counterexamples appear to presuppose a broader notion of “undermining defeat” than that given here using the subjunctive conditional, though clearly the question of how well the suggestion could resist his objections would require careful work and can’t be answered here.

The second reason is that I fear that the characterization of an undermining defeater is too strong. Consider again Orange:

ORANGE

$E_O$  That wall appears to be orange.

$P_O$  That wall is orange.

$R_O$  That wall is not yellow.

Since  $P_O$  entails  $R_O$ ,  $\neg R_O$  is evidence against  $P_O$ . But in a context where one is staring at an orange-looking wall, it seems to me that any evidence one might gain that the wall is not orange would also be evidence against the claim that if the wall were not orange, it would not appear to be orange. “Well, it clearly *looks* orange, but now I’ve got evidence that it’s *not* orange, so that leads me to doubt that it wouldn’t look orange unless it were orange.” So I fear that in this case  $\neg R_O$  could turn out to satisfy the characterization of an undermining defeater. Obviously this would spell trouble for the proposal about transmission failure, since transmission seems clearly not to fail in Orange. More seriously, though, it makes the Pollock-inspired characterization of an undermining defeater look suspect.

## References

- Brown, J. (2004). Wright on transmission failure. *Analysis*, 64(1), 57–67.
- Chandler, J. (2010). The transmission of support: A Bayesian re-analysis. *Synthese*, 176(3), 333–343.
- Cohen, S. (2002). Basic knowledge and the problem of easy knowledge. *Philosophy and Phenomenological Research*, 65(2), 309–329.
- Cohen, S. (2005). Why basic knowledge is easy knowledge. *Philosophy and Phenomenological Research*, 70(2), 417–430.
- Davies, M. (2003). Externalism, self-knowledge, and transmission of warrant. In M. Frápolli & E. Romero (Eds.), *Meaning, basic self-knowledge, and mind: Essays on Tyler Burge* (pp. 105–130). CSLI Publications: Stanford.
- Davies, M. (2004). Epistemic entitlement, warrant transmission, and easy knowledge. *Aristotelian Society Supplement*, 78, 213–245.
- DeRose, K. (1994). Lewis on ‘might’ and ‘would’ counterfactuals. *Canadian Journal of Philosophy*, 24, 413–418.
- Dretske, F. I. (1970). Epistemic operators. *The Journal of Philosophy*, 67(24), 1007–1023.
- Hawthorne, J. (2004). The case for closure. In M. Steup, (Ed.), *Contemporary Debates in Epistemology*. Blackwell.
- Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell Publishing.
- Neta, R. (2007). Fixing the transmission: The new Mooreans. In S. Nuccetelli & G. Seay (Eds.), *Themes from G. E. Moore: New essays in epistemology and ethics*. Oxford: Oxford University Press.
- Okasha, S. (2004). Wright on the transmission of support: A Bayesian analysis. *Analysis*, 64(2), 139–146.
- Pollock, J. (1974). *Knowledge and justification*. Princeton, NJ: Princeton University Press.
- Pryor, J. (2004). What’s wrong with Moore’s argument? *Philosophical Issues*, 14(1), 349–378.
- Pryor, J. (2009). When warrant transmits. In A. Coliva (Ed.), *Wittgenstein, epistemology and mind: Themes from the philosophy of Crispin Wright*. Oxford: Oxford University Press.
- Silins, N. (2005). Transmission failure failure. *Philosophical Studies*, 126(1), 71–102.
- Silins, N. (2007). Basic justification and the Moorean response to the skeptic. In T. Gendler & J. Hawthorne (Eds.), *Oxford studies in epistemology* (Vol. II). Oxford: Oxford University Press.
- Smith, M. (2009). Transmission failure explained. *Philosophy and Phenomenological Research*, 79(1), 164–189.
- Stalnaker, R. (1984). *Inquiry*. Cambridge, MA: MIT Press.
- White, R. (2006). Problems for dogmatism. *Philosophical Studies*, 131(3), 525–557.
- Wright, C. (1985). Facts and certainty. *Proceedings of the British Academy*, 71, 429–472.
- Wright, C. (2000). Cogency and question-begging: Some reflections on McKinsey’s paradox and putnam’s proof. *Noûs*, 34, 140–163.
- Wright, C. (2002). (Anti-)sceptics simple and subtle: G. E. Moore and John McDowell. *Philosophy and Phenomenological Research*, 65(2), 330–348.
- Wright, C. (2003). *Some reflections on the acquisition of warrant by inference* (pp. 57–77). Cambridge, MA: MIT Press.
- Wright, C. (2008). The perils of dogmatism. In S. Nuccetelli & G. Seay (Eds.), *Themes from G. E. Moore: New essays in epistemology and ethics* (pp. 25–48). Oxford: Oxford University Press.