# The Bayesian Explanation of Transmission Failure

Geoff Pynn

May 13, 2009

**Abstract**

Transmission failure occurs when a subject's $e$-based justification for $p$ cannot provide her with justification to believe $q$, despite the fact that $q$ is one of $p$'s known entailments. According to the Bayesian explanation, transmission failure occurs because $P(q|e) < P(q)$. I argue that the Bayesian explanation is extensionally inadequate: the condition it identifies is neither necessary nor sufficient for transmission failure.

## 1 Transmission failure

Typically, when you gain justification for believing $p$ and recognize that $p$ entails $q$, you thereby also gain justification for believing $q$. But in some cases, your justification for believing $p$ can "fail to transmit" to $q$; i.e., recognizing that $q$ is entailed by your justified belief in $p$ is *not* a way of gaining justification to believe $q$. [1]

Let us distinguish the idea that justification is *closed under known entailment* from the idea that justification *always transmits across known entailment*. Here is a statement of the closure principle:

> **Closure.** If $S$ has justification to believe $p$, and recognizes that $p$ entails $q$, then $S$ has justification to believe $q$.

[1]The term 'transmission failure' was coined by Crispin Wright, see Wright 1985, Wright 2000, Wright 2002, and Wright 2008. See also Neta 2007,Okasha 2004, Pryor 2004, Pryor 2009, Silins 2005, and Silins 2007 for relevant recent discussions. Some, such as Silins 2005, deny the existence of transmission failure. Here I shall assume that transmission failure is a real phenomenon.

Closure gives a sufficient condition for *having* justification to believe $q$. Let us assume that Closure is correct. This assumption does not imply that Transmission is exceptionless:

> **Transmission.** If $S$ acquires justification to believe $p$, then, provided $S$ recognizes that $p$ entails $q$, $S$ thereby also acquires justification to believe $q$.

Suppose that $S$ gains evidence $e$, which gives her justification to believe $p$. If she does *not* thereby also acquire justification to believe $q$, which she recognizes to be entailed by $p$, then we have a counterexample to Transmission; i.e., an instance of *transmission failure*.

Examples of transmission failure are widely recognized. For example:

***RED***

$e$ That wall looks red.

$p$ That wall is red.

$q$ That wall is not white and illuminated by red lights. [2]

Suppose that getting evidence $e$ gives you justification to believe $p$. If this is the case, then merely recognizing that $p$ entails $q$ is not a way of acquiring justification to believe $q$. [3]

Those who accept Closure agree that, in *RED*, a subject has justification to believe $p$ only if she also has justification to believe $q$. But if they accept that *RED* is a case of transmission failure, they will insist that this justification must come from some other source than her $e$-based justification for $p$. There are a number of candidate sources. She might know that all of the light bulbs in the house she is in are white, or that, generally speaking, it is quite rare for a white wall to be convincingly illuminated by red light. But her $e$-based justification for $p$ is not a candidate source of justification for $q$.

What is it about an $e$-$p$-$q$ triad that makes it an instance of transmission failure? [4]

[2] This example is originally due to Dretske 1970 and has been discussed more recently by Cohen 2002, Pryor 2004, Wright 2002, Wright 2008 and others.

[3] Somewhat incredibly, James Pryor appears to deny this. He writes that your $e$-based justification to believe $p$ "contributes to the credibility of the claim that the wall isn't white and lit by tricky lights" (Pryor 2004, 362).

[4] Much of the literature on transmission failure has focused on whether and how the so-called 'McKinsey paradox' and G. E. Moore's 'proof of the external world' are instances of

A recent answer to this question is the *Bayesian explanation* of transmission failure. [5] I will first describe the Bayesian explanation, and then offer a number of counterexamples that show that it is extensionally inadequate. After considering—and rejecting—several possible lines of response on behalf of the Bayesian explanation, I will draw two brief morals.

## 2 The Bayesian explanation

According to the Bayesian explanation, transmission failure occurs when the probability of $q$ conditional on $e$ is lower than the unconditional probability of $q$; i.e., when $P(q|e) < P(q)$. Plausible assumptions imply that in ***RED*** this inequality is satisfied.

In ***RED***, $\neg q$ entails $e$; i.e., if the wall *is* white and illuminated by red lights, it looks red. [6] So

$$P(e|\neg q) \approx 1$$

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

$$P(e) < 1$$

Hence $P(e|\neg q) > P(e)$. By Bayes' Theorem, $P(\neg q|e) > P(\neg q)$. [7] So

$$P(q|e) < P(q)$$

Why should satisfaction of this inequality lead to transmission failure? We understand $P$ as a function from propositions to either the credences or evidential probabilities of a subject. Given either interpretation, when $P(p|e) < P(p)$ the subject ought rationally to lower her confidence in $p$ on learning $e$. Thus in

---

transmission failure. Wright 2000 and his respondents are concerned with the McKinsey paradox;Davies 2009, Pryor 2004, and Silins 2007 with Moore's proof. But transmission failure is an interesting and puzzling phenomenon in its own right, regardless of its utility for solving other philosophical problems.

[5] See Silins 2007 for the most recent version of the Bayesian explanation. Hawthorne 2004a, fn. 42, Okasha 2004, Cohen 2005, 424-425, and White 2006 all make similar suggestions.

[6] Properly speaking, $\neg q$ does not *entail* $p$; if there is a film of yellow plastic wrap covering the wall, it looks 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 have any hope at all. Troubled readers can replace *entails* with something along the lines of: *all but guarantees, provided relevant background details are held fixed.*

[7] I assume throughout that $P(\neg q) > 0$. The assumption is apt: it was not certain in advance that the wall would not be white and illuminated by red lights.

*RED* she ought rationally to lower her confidence in $q$ on learning $e$. But how could she *gain* justification to believe $q$ by acquiring evidence that rationally enjoins her to *lower* her confidence in $q$? Acquiring $e$ couldn't *directly* justify $q$, since the rational response to learning $e$ is to become less confident in $q$. And $e$ could not furnish her with *indirect* justification for $q$ merely by giving her justification $p$. If $e$ does not even give her indirect justification to believe $q$, then her $e$-based justification for $p$ fails to transmit to $q$. This is the Bayesian explanation. [8]

Here is one potential difficulty with the Bayesian explanation as applied to *RED*. Suppose that when I have a visual experience with the content $p$, I thereby learn $p$. And suppose that when the wall looks red to me I have a visual experience with the content *that wall is red*. Then, since in *RED* $p$ entails $q$, $P(q|p) = 1$. Thus when the wall looks red to me, I learn that the wall is red, and thereby ought rationally to set my credence in $q$ to 1. 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 $q$ to go down, but rather to go up.

But, one might think, so much the worse for this analysis of the epistemology of visual experience. In the next section, I shall consider more general difficulties for the Bayesian explanation of transmission failure.

# 3 Problem cases

Key to the Bayesian explanation for what goes wrong in *RED* is that $\neg q$ entails $e$. Provided that $P(e) < 1$, this entailment guarantees that $P(q|e) < P(q)$. But it is inessential to transmission failure that $\neg q$ entails $e$. When $\neg q$ does not entail $e$, there is no guarantee that the problematic inequality is satisfied. Two kinds of cases demonstrate that $P(q|e) < P(q)$ is unnecessary to transmission failure.

[8] 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 (¬BIV) is an instance of transmission failure, since, as he puts it: "[I]f Moore's experience makes his justification to believe ¬BIV go up, then it can't be that his confidence in ¬BIV should instead go down." Contraposing: if his confidence in ¬BIV should go down, his experience does not make his justification to believe ¬BIV go up.

## 3.1 Transmission failure when $P(q|e) \approx P(q)$

When $e$ and $\neg q$ are independent, $P(\neg q|e) \approx P(\neg q)$, so $P(q|e) \approx P(q)$, and hence it is not the case that $P(q|e) < P(q)$. For example, suppose it is 6:05 in the morning:

***BUS***

$e$ Meg's train was scheduled to arrive at six o'clock.

$p$ Meg will arrive shortly.

$q$ Meg wasn't just killed by a bus crossing the street.

*BUS* is a case of transmission failure: if $e$ is what gives you justification for $p$, you cannot acquire justification for $q$ merely by recognizing that it is entailed by $p$. But Meg's being killed by a bus crossing the street is unrelated to her train's schedule. So $P(\neg q|e) \approx P(\neg q)$, and hence $P(q|e) \approx P(q)$. Thus the Bayesian explanation does not predict that *BUS* is a case of transmission failure.

It is very easy to construct cases like *BUS*. First, take some evidence $e$ that is compatible with $\neg p$ but is ordinarily regarded as sufficient justification for $p$. Second, imagine an improbable scenario $s$ in which (i) $\neg p$ but (ii) it is quite likely that a subject would believe $p$ on the basis of $e$. Let $q = \neg s$. The resulting $e$-$p$-$q$ triad will be a case of transmission failure. Clearly it is not necessary in such a scenario that $\neg q$ entail $e$.

## 3.2 Transmission failure when $P(q|e) > P(q)$

*BUS*-like cases can be captured with a modest tweak to the Bayesian account. We simply replace the "$<$" with "$\lessapprox$": an $e$-$p$-$q$ triad is a case of transmission failure just in case $P(q|e) \lessapprox P(q)$. *BUS* satisfies this condition. However, we can also construct cases of transmission failure where $e$ *raises* the probability of $q$. Consider:

***RED****

$e$ That wall looks red.

$p$ That wall is red.

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

*RED** is a case of transmission failure. Just as in *BUS*, $\neg q$ does not entail $e$—if the wall is white and lit by colored lights it might look green, not red.

Now add the following background knowledge:

> You are in a house with 100 rooms. 51 are painted white, and 49 are painted red. Of the white rooms, three are lit by green lights and one is lit by a red light. All of the rest of the rooms are lit by white lights. You're about to open the door to the first room in the house.

We can now assign the following values. Since 50 of the 100 rooms look red (the 49 red rooms, and the one white room lit by red lights),

$$P(e) = 0.5$$

Since, of the four white rooms lit by colored lights, one looks red,

$$P(e|\neg q) = 0.25$$

Thus $P(e|\neg q) < P(e)$. By Bayes' Theorem, $P(\neg q|e) < P(\neg q)$, and hence $P(q|e) > P(q)$. So the modified Bayesian explanation fails to predict that ***RED**** is a case of transmission failure. Yet this is, if anything, an even clearer case than ***RED***. Since your background knowledge indicates that one of the white rooms *is* illuminated by red lights, you would seem *particularly* unjustified in concluding $q$ from your $e$-based belief that $p$.

Constructing ***RED****-like cases is straightforward. Start with an $e$-$p$-$q$ triad that is a clear-cut case of transmission failure where $\neg q$ entails $e$. Then strengthen $q$ in such a way that preserves the intuition of transmission failure. By strengthening $q$, you make it the case that $\neg q$ no longer entails $e$, and the problematic inequality is no longer satisfied (or at least, it is not clearly satisfied). [9] Then supply the relevant background information so that $P(q|e) > P(e)$. This procedure will result in a case of transmission failure left behind by the Bayesian explanation.

---

[9]For example, consider:

> ***ZEBRA***
>
> $e$ That animal is horse-shaped and has black-and-white stripes.
>
> $p$ That animal is a zebra.
>
> $q$ That animal is not a mule cleverly painted to look like a zebra.

*ZEBRA* satisfies the problematic inequality. If we replace $q$ with $r$, the intuition of transmission failure remains, but the problematic inequality vanishes:

> $r$ That animal is not a mule cleverly painted to look like a different kind of animal.

The Bayesian explanation predicts that transmission fails from $p$ to $q$, but not from $p$ to $r$.

## 3.3 Bayesian responses

The proponent of the Bayesian explanation could respond to these cases in one of two ways. The first is to deny that *BUS* and *RED** are cases of transmission failure. The second is to more drastically modify the Bayesian explanation to accommodate *BUS* and *RED**.

The first response is not appealing; if *RED* is a case of transmission failure, so are *BUS* and *RED**.

The second response would be as plausible as whatever modification accompanies it. I will consider two candidates.

Notice that if the wall is not white and lit by colored lights, then it is not white and lit by red lights; i.e., *RED**'s $q$ entails *RED*'s $q$.. Perhaps the transmission failure exhibited by *RED* somehow "infects" *RED**. How could we modify the Bayesian explanation to capture this infection? Here is the obvious way to do so. Let us suppose that an $e$-$p$-$q$ triad is an instance of transmission failure just in case **either** $P(q|e) \lessapprox P(q)$ (ordinary cases) **or** $q$ entails some proposition $r$ such that $P(r|e) \lessapprox P(r)$ ("infection" cases). This account would accurately predict that *RED** is a case of transmission failure.

But the proposal is doomed, for it is far too strong. Virtually any case of transmission will count as an "infection" case. Consider:

***ORANGE***

$e$ That wall looks orange.

$p$ That wall is orange.

$q$ That wall is not yellow.

*ORANGE* is clearly not a case of transmission failure. However, *ORANGE*'s $q$ entails

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

By the same reasoning as we employed regarding *RED*, $P(r|e) < P(r)$. Thus the proposed modification implies that *ORANGE* is a case of transmission failure.

In a very interesting manuscript, Matt Kotzen argues that when $e$ confirms $p$, $e$ also confirms $p$'s entailment $q$ only if $P(q) < P(p|e)$. [10] 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 $P(q|e) < P(q)$, the Dragging Condition *can't* be satisfied, so any case that exhibits the

[10] Kotzen ms.

problematic inequality at the heart of the Bayesian explanation will also fail to satisfy the Dragging Condition. [11] But a triad could fail to satisfy the Dragging Condition even if $P(q|e) > P(q)$. Perhaps, then, the Dragging Condition will enable us to craft an account that captures what is wrong in *BUS* and *RED** without being overly strong. Following Kotzen, let us propose that an $e$-$p$-$q$ triad is a case of transmission failure just in case $P(p|e) \leq P(q)$.

This second modification is promising with respect to *BUS*. Since, in that case, $P(q|e) \approx P(q)$, given that $p$ entails $q$, $P(p|e) \leq P(q|e)$. Hence $P(p|e) \lessapprox P(q)$. So in *BUS* the Dragging Condition is not satisfied.

Unfortunately, Kotzen's proposal fails to predict that *RED** is a case of transmission failure. In *RED**, you know that 96 of 100 rooms are not white and lit by colored lights, so

$$P(q) = 0.96$$

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

$$P(p|e) = 0.98$$

Thus $P(q) < P(p|e)$; the Dragging Condition is satisfied. So the proposal is too weak. [12]

There may be another way to modify the Bayesian explanation to account for the fact that *RED** is a case of transmission failure. But it is not obvious what that way would be.

# 4 Does the Bayesian explanation yield a sufficient condition for transmission failure?

The previous section shows that $P(q|e) < P(q)$ is unnecessary for transmission failure. However, the proponent of the Bayesian explanation might suggest that satisfaction of this inequality is a *sufficient* condition for transmission failure. In this section I shall raise doubts about this suggestion. I'll argue:

1. If there are cases where $e$ gives you justification for $p$, but $P(p|e) < P(p)$, then there is little reason to suppose that in an $e$-$p$-$q$ triad $e$ could not transmit justification to $q$ if $P(q|e) < P(q)$.

[11] Since $p$ entails $q$, $P(p|e) \leq P(q|e)$. Thus, if $P(q|e) < P(q)$, $P(p|e) < P(q)$.

[12] This is no complaint against the Dragging Condition itself, which I find to be quite plausible as concerns the technical notion of confirmation. Evidence for $p$ can confirm one of $p$'s entailments without transmitting justification to that entailment.

2. There are cases where $e$ gives you justification for $p$, but $P(p|e) < P(p)$.

3. So, there is little reason to suppose that the Bayesian explanation yields a sufficient condition for transmission failure.

Let me first say why the second premise is correct. Recall the reasoning at the heart of the Bayesian explanation. We assume that when $P(q|e) < P(q)$, one ought rationally to lower one's confidence in $p$ upon learning $e$. And if the rational response to learning $e$ is to *lower* your confidence in $q$, then it seems that $e$ couldn't *increase* your justification to believe $q$. So when $P(q|e) < P(q)$, we have a case of transmission failure. However, if there *are* cases where some evidence $e$ gives you justification to believe $p$ despite the fact that $P(p|e) < P(p)$, this line of reasoning is severely undermined.

Now, to the second premise: there seem to be such cases.

The first is borrowed from Jonathan Vogel. [13] You are at the Wealth and Privelege Invitational Golf Tournament, which has a notoriously difficult hole known as the 'Heartbreaker'. Crazy Carl offers you an even-money bet that all sixty players will get a hole-in-one on the Heartbreaker. You reason to yourself:

$e$ The odds that every player will get a hole-in-one on the Heartbreaker are ludicrously slim.

$p$ Crazy Carl won't beat those ludicrously slim odds.

Clearly, $e$ epistemically supports belief in $p$. But if Crazy Carl *will* beat those ludicrously slim odds, the odds are still ludicrously slim, so $\neg p$ entails $e$. Thus $P(e|\neg p) = 1$. The odds might not have been so ludicrously slim; for example, the players might all have had homing devices in their balls. So $P(e) < 1$. Thus $P(e|\neg p) > P(e)$, and by Bayes' Theorem $P(\neg p|e) > P(\neg p)$. Thus $P(p|e) < P(p)$.

Here is a second case. Suppose we are sitting in an air-conditioned hotel room in the middle of downtown Phoenix on July 21. I ask, jokingly, "Do you think it will snow here today?" You look down at the newspaper and reply: "It's July 21 in Phoenix; of course it's not going to snow here today." Your response is perfectly in order. You have cited $e$ in support of $p$:

$e$ It's July 21 in Phoenix.

$p$ It's not going to snow here today.

[13] Vogel 1999.

The negation of $p$ would be expressed by the sentence, "It is going to snow here today." Uttered or thought in the same context as $p$, that sentence's content entails that it is July 21 in Phoenix. Thus $\neg p$ entails $e$, so $P(e|\neg p) = 1$. Of course it might not have been July 21 in Phoenix, so $P(e) < 1$. Thus $P(e|\neg p) > P(e)$; by Bayes' theorem $P(\neg p|e) > P(\neg p)$; and so $P(p|e) < P(p)$.

These cases are prima facie examples of evidence $e$ that furnishes justification for $p$ despite the fact that $P(p|e) < P(p)$. One could cling dogmatically to the view that if $e$ and $p$ are so related, $e$ could not possibly yield justification for $p$. Such dogmatic clinging would require re-describing the foregoing cases in such a way that either the relevant $e$ does not confer justification on $p$, or that it's not the case that $P(p|e) < P(p)$. Without such re-description, the cases provide clear support for the second premise above.

Accordingly, we have little reason to think that whenever $P(q|e) < P(q)$, an $e$-$p$-$q$ triad is an instance of transmission failure.

# 5 Two brief morals

I have argued that the Bayesian explanation posits a condition which is neither necessary nor sufficient for transmission failure. I shall close with two brief morals.

First moral. Transmission failure is a phenomenon that concerns evidence, justification, and belief. The Bayesian explanation attempts to account for this phenomenon in terms of a particular inequality between subjective or evidential probabilities. But the presence of this inequality does not correspond to the phenomenon of transmission failure. Nor does the presence of two other candidate inequalities which we considered in section 3.3. One simple explanation for this divergence is that ordinary justification and belief cannot be accurately modeled by relations among subjective or evidential probabilities. We can add the failure of the Bayesian explanation to our list of reasons to resist characterizing ordinary epistemic notions in probabilistic terms.

Second moral. Each of the examples of transmission failure I have considered involves what John Hawthorne calls a "lottery proposition": in each case, $q$ is a proposition which, though highly likely, we are inclined to think *can't be known* on the basis of $e$. [14] Equally, in each case, $p$ is a proposition

[14] Hawthorne 2004b. Hawthorne's notion of a lottery proposition does not involve a relation to particular evidence, but it should, since our intuitive judgments about when a proposition is or isn't known involve assumptions about the sorts of evidence one has for those propositions.

we are inclined to think *can be known* on the basis of $e$, at least in ordinary circumstances in which we are not too worried about skeptical possibilities. Though Transmission concerns justification and not knowledge, consider this modification:

> **Transmission-K.** If $S$ acquires knowledge-level justification for $p$, then, provided $S$ recognizes that $p$ entails $q$, $S$ thereby also acquires knowledge-level justification to believe $q$.

On the assumption that $e$ yields knowledge-level justification for $p$—which seems intuitively plausible in each of the cases discussed above—Transmission-K would imply that it also yields knowledge-level justification for $q$. But given that $q$ is a lottery proposition with respect to $e$, this is intuitively incorrect. If our inclination to regard the cases above as instances of transmission failure reflects our intuitive sense that $q$ can't be known on the basis of $e$, then the place to look for an account of transmission failure is in an account of lottery propositions. The correct explanation of what makes for a lottery proposition might lead us to a satisfying account of transmission failure.

## References


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. (2009). Two purposes of arguing and two epistemic projects. In Ravenscroft, I., editor, *Frank Jackson and His Critics*. Blackwell.

Dretske, F. I. (1970). Epistemic operators. *The Journal of Philosophy*, 67(24):1007–1023.

Hawthorne, J. (2004a). The case for closure. In Steup, M., editor, *Contemporary Debates in Epistemology*. Blackwell.

Hawthorne, J. (2004b). *Knowledge and Lotteries*. Oxford University Press.

Kotzen, M. (ms). Dragging and confirming.

Neta, R. (2007). Fixing the transmission: The new mooreans. In Nuccetelli, S. and Seay, G., editors, *Themes from G. E. Moore: New Essays in Epistemology and Ethics*. Oxford University Press.

Okasha, S. (2004). Wright on the transmission of support: a bayesian analysis. *Analysis*, 64(2):p139 – 146.

Pryor, J. (2004). What's wrong with moore's argument? *Philosophical Issues*, 14(1):349–378.

Pryor, J. (2009). When warrant transmits. In Coliva, A., editor, *Wittgenstein, Epistemology and Mind: Themes from the Philosophy of Crispin Wright*. Oxford University Press.

Silins, N. (2005). Transmission failure failure. *Philosophical Studies*, (126):71–102.

Silins, N. (2007). Basic justification and the moorean response to the skeptic. In Gendler, T. and Hawthorne, J., editors, *Oxford Studies in Epistemology*, volume II. Oxford University Press.

Vogel, J. (1999). The new relevant alternatives theory. *Philosophical Perspectives*, 13:155–180.

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. (2008). The perils of dogmatism. In Nuccetelli, S. and Seay, G., editors, *Themes from G. E. Moore: New Essays in Epistemology and Ethics*, pages 25–48. Oxford University Press.