The rational response to unexpected news from a reliable source

Eric Johannesson

October 14, 2023

Abstract

When a source you believe to be reliable contradicts a belief of yours, you seem to be faced with a dilemma: should you weaken your belief, or should you weaken your belief in the reliability of the source? We show that, under very reasonable assumptions, this is no dilemma at all: you should do both.

With the notable exceptions of Jon Voight, James Woods, and Clint Eastwood, most Hollywood celebrities did not endorse Donald Trump during the US presidential election of 2016. Prior to his endorsement of Trump, I held Clint Eastwood in quite high regard. I liked his movies (still do), and I thought he seemed like a decent guy. At least for me, part of holding someone in high regard is to consider him to be a good judge of character. I was now faced with a dilemma (or so it seemed): should I lower my regard for Clint Eastwood (whom I initially admired), or should I increase my regard for Donald Trump (whom I initially loathed)?

Then it dawned on me that perhaps this was no dilemma at all. Rationally speaking, unless I were completely certain initially that Trump was a bad guy, or that Clint Eastwood was a good judge of character – and I do not think anyone should be completely certain of such things – I would be required to do both things simultaneously. This may sound paradoxical: how can Eastwood's endorsement of Trump be evidence that Trump is a good guy, while also be evidence that Eastwood is a poor judge of character? But it actually follows given very reasonable assumptions.

To see why, let G be the proposition that Clint Eastwood is a good judge of character, H that Donald Trump is a bad guy, and E that Clint Eastwood endorses Donald Trump. Let us assume, by way of specifying what I mean by 'a good judge of character', that

(1) $\quad P(E|G \wedge H) < P(E|\neg G \wedge H) = P(E|\neg G \wedge \neg H) < P(E|G \wedge \neg H).$

Clearly, in order for these conditional probabilities to be defined in the first place, we need to assume that

(2) $\quad 0 < P(G) < 1$ and $0 < P(H) < 1$.

In any case, I think this assumption is reasonable. Let us also assume, partly by way of simplification, but arguably also as a reflection of my initial beliefs at the time, that

(3) $\quad P(G \wedge H) = P(G)P(H).$

In the appendix below, we show that these three assumptions non-trivially[1] imply the following conclusion:

(4) a. $P(H|E) < P(H)$, and
b. there is a threshold $0 < r < 1$ given by

$$r = \frac{P(E|G \wedge \neg H) - P(E|\neg G \wedge H)}{P(E|G \wedge \neg H) - P(E|G \wedge H)}$$

such that

$$P(G|E) < P(G) \text{ if } P(H) > r,$$
$$P(G|E) > P(G) \text{ if } P(H) < r, \text{and}$$
$$P(G|E) = P(G) \text{ if } P(H) = r,$$

with $r = 1/2$ in the special case of

$$P(E|\neg G \wedge H) = \frac{P(E|G \wedge H) + P(E|G \wedge \neg H)}{2}$$

In other words, my belief that Donald Trump is a bad guy weakens. Furthermore, provided that this belief was sufficiently strong in the first place, my belief that Clint Eastwood is a good judge of character also weakens. For example, with $P(G) = P(H) = 0.9$, $P(E|G \wedge H) = 0.1$, $P(E|G \wedge \neg H) = 0.9$ and $P(E|\neg G \wedge H) = P(E|\neg G \wedge \neg H) = 0.5$, we get $P(G|E) \approx 0.76$ and $P(H|E) \approx 0.59$.

These results obviously apply to cases of unexpected criticism as well. Suppose, for instance, that you have written a paper that you think is good enough to be published in a good journal. Let E be the proposition that the journal rejects your paper, G that the journal is a good judge of quality, and H that your paper is excellent. Under assumptions (1)-(3) above (which still seem reasonable, unless you are an editor of the journal, in which case (3) seems dubious), the rejection of your paper is evidence that neither the paper nor the journal are as good as you initially thought they were.[2]

To accept these conclusions (with our without proof), I think it helps to point out that, when unexpected things happen, we must question all the beliefs that lead us not to expect it. Two beliefs lead me to expect that Clint Eastwood would not endorse Donald Trump: the belief that Clint Eastwood is a good judge of character, and the belief that Donald Trump is a bad guy. Given the unexpected endorsement, it only seems fair that I should question both of these beliefs. The appearance of paradox is due to the fact that, by questioning the first belief, I am effectively questioning my reason for thinking that the evidence has any negative bearing on the second belief, and vice versa. The lesson, I take it, is that this feedback loop reaches an equilibrium where both beliefs are weakened.

[1]That is, in the sense that the assumptions are consistent with the laws of probability, so not implying just anything.

[2]The result also implies that, if your paper instead were to be accepted, that would increase your confidence in the quality of both the paper and the journal. So the result actually applies to all cases of praise and criticism, unexpected or not.

Appendix

To show that the assumptions imply the conclusion *non-trivially*, we need to make sure that they are consistent with the laws of probability in the first place. If they are not, the conclusion will not be that interesting. Intuitively, they should be (I just described a situation where they seem to hold), but one can never be too careful about these things. To proceed with the formal argument, we will assume that the propositions E, G, H are logically independent. To establish the consistency of our assumptions, it is sufficient to establish that they are satisfied by a probability distribution on the eight mutually exclusive propositions $(E \wedge G \wedge H), (E \wedge G \wedge \neg H), ..., (\neg E \wedge \neg G \wedge \neg H)$. It suffices to make sure that, for any choice of real numbers $0 \leq a, b, c, g, h \leq 1$, there is a probability distribution P on these eight propositions such that $P(G) = g$, $P(H) = h$, $P(E|G \wedge H) = a$, $P(E|G \wedge \neg H) = b$, $P(E|\neg G \wedge H) = P(E|\neg G \wedge \neg H) = c$, and $P(G \wedge H) = gh$. By the laws of probability and our assumptions, we see that any such choice actually determines an assignment of real numbers to the eight propositions, as follows:[3]

$$\begin{array}{lclcl}
P(E \wedge G \wedge H) & = & P(E|G \wedge H)P(G \wedge H) & = & agh \\
P(E \wedge G \wedge \neg H) & = & P(E|G \wedge \neg H)P(G \wedge \neg H) & = & bg(1-h) \\
P(E \wedge \neg G \wedge H) & = & P(E|\neg G \wedge H)P(\neg G \wedge H) & = & c(1-g)h \\
P(E \wedge \neg G \wedge \neg H) & = & P(E|\neg G \wedge \neg H)P(\neg G \wedge \neg H) & = & c(1-g)(1-h) \\
P(\neg E \wedge G \wedge H) & = & P(\neg E|G \wedge H)P(G \wedge H) & = & (1-a)gh \\
P(\neg E \wedge G \wedge \neg H) & = & P(\neg E|G \wedge \neg H)P(G \wedge \neg H) & = & (1-b)g(1-h) \\
P(\neg E \wedge \neg G \wedge H) & = & P(\neg E|\neg G \wedge H)P(\neg G \wedge H) & = & (1-c)(1-g)h \\
P(\neg E \wedge \neg G \wedge \neg H) & = & P(\neg E|\neg G \wedge \neg H)P(\neg G \wedge \neg H) & = & (1-c)(1-g)(1-h)
\end{array}$$

We verify that this is an assignment of probabilities. Clearly, each of the eight propositions is assigned a real number between 0 and 1. Their sum is given by

$$\begin{aligned}
&[ag + c(1-g) + (1-a)g + (1-c)(1-g)]h \\
+&[bg + c(1-g) + (1-b)g + (1-c)(1-g)](1-h) \\
=&[g + (1-g)]h + [g + (1-g)](1-h) = h + (1-h) = 1
\end{aligned}$$

[3]Given the laws of probability, assumption (3) implies

$$\begin{aligned}
P(G \wedge \neg H) &= P(G) - P(G \wedge H) = P(G)(1 - P(H)) = P(G)P(\neg H) \\
P(\neg G \wedge H) &= P(H) - P(G \wedge H) = (1 - P(G))P(H) = P(\neg G)P(\neg H) \\
P(\neg G \wedge \neg H) &= 1 - (P(G) - P(G \wedge H)) - (P(H) - P(G \wedge H)) = P(\neg G)P(\neg H)
\end{aligned}$$

We also use the general fact that, for any propositions X and Y with $P(Y) > 0$, we have

$$P(\neg X|Y) = \frac{P(\neg X \wedge Y)}{P(Y)} = \frac{P(Y) - P(X \wedge Y)}{P(Y)} = 1 - P(X|Y)$$

In turn, we now have:

$$\begin{aligned}P(G) &= P(E \wedge G \wedge H) + P(E \wedge G \wedge \neg H) + P(\neg E \wedge G \wedge H) + P(\neg E \wedge G \wedge \neg H)\\ &= g(ah + b(1-h) + (1-a)h + (1-b)(1-h))\\ &= g(h + (1-h)) = g\end{aligned}$$

$$\begin{aligned}P(H) &= P(E \wedge G \wedge H) + P(E \wedge \neg G \wedge H) + P(\neg E \wedge G \wedge H) + P(\neg E \wedge \neg G \wedge H)\\ &= h(ag + c(1-g) + (1-a)g + (1-c)(1-g))\\ &= h(h + (1-h)) = h\end{aligned}$$

$$\begin{aligned}P(E|G \wedge H) &= \frac{P(E \wedge G \wedge H)}{P(G \wedge H)} = \frac{P(E \wedge G \wedge H)}{P(E \wedge G \wedge H) + P(\neg E \wedge G \wedge H)}\\ &= \frac{agh}{agh + (1-a)gh} = a\end{aligned}$$

$$\begin{aligned}P(E|G \wedge \neg H) &= \frac{P(E \wedge G \wedge \neg H)}{P(G \wedge \neg H)} = \frac{P(E \wedge G \wedge \neg H)}{P(E \wedge G \wedge \neg H) + P(\neg E \wedge G \wedge \neg H)}\\ &= \frac{bg(1-h)}{bg(1-h) + (1-b)g(1-h)} = b\end{aligned}$$

$$\begin{aligned}P(E|\neg G \wedge H) &= \frac{P(E \wedge \neg G \wedge H)}{P(\neg G \wedge H)} = \frac{P(E \wedge \neg G \wedge H)}{P(E \wedge \neg G \wedge H) + P(\neg E \wedge \neg G \wedge H)}\\ &= \frac{c(1-g)h}{c(1-g)h + (1-c)(1-g)h} = c\end{aligned}$$

$$\begin{aligned}P(E|\neg G \wedge \neg H) &= \frac{P(E \wedge \neg G \wedge \neg H)}{P(\neg G \wedge \neg H)} = \frac{P(E \wedge \neg G \wedge \neg H)}{P(E \wedge \neg G \wedge \neg H) + P(\neg E \wedge \neg G \wedge \neg H)}\\ &= \frac{c(1-g)(1-h)}{c(1-g)(1-h) + (1-c)(1-g)(1-h)} = c\end{aligned}$$

$$\begin{aligned}P(G \wedge H) &= P(E \wedge G \wedge H) + P(\neg E \wedge G \wedge H)\\ &= agh + (a-1)gh = gh\end{aligned}$$

Hence, we have established consistency. We will now establish the conclusion. Let a, b, c, g, h be as before, and assume that they also satisfy our assumptions, which is to say that $0 < a < c < b < 1$ and $0 < g, h < 1$. We observe that

$$\begin{aligned}P(G|E) &= \frac{P(E \wedge G)}{P(E)} = \frac{P(E \wedge G \wedge H) + P(E \wedge G \wedge \neg H)}{P(E)}\\ &= \frac{agh + bg(1-h)}{P(E)} = P(G)\frac{ah + b(1-h)}{P(E)}\end{aligned}$$

and

$$\begin{aligned}P(H|E) &= \frac{P(E \wedge H)}{P(E)} = \frac{P(E \wedge G \wedge H) + P(E \wedge \neg G \wedge H)}{P(E)} \\ &= \frac{agh + c(1-g)h}{P(E)} = P(H)\frac{ag + c(1-g)}{P(E)}\end{aligned}$$

Moreover, we have

$$\begin{aligned}P(E) =& P(E \wedge G \wedge H) + P(E \wedge G \wedge \neg H) + P(E \wedge \neg G \wedge H) + P(E \wedge \neg G \wedge \neg H) \\ =& agh + bg(1-h) + c(1-g)h + c(1-g)(1-h) \\ =& agh + bg(1-h) + c(1-g)\end{aligned}$$

Putting these things together, we obtain

$$P(G|E) = P(G)\frac{ah + b(1-h)}{agh + bg(1-h) + c(1-g)}$$

and

$$P(H|E) = P(H)\frac{ag + c(1-g)}{agh + bg(1-h) + c(1-g)}$$

Thus, have $P(H|E) < P(H)$ just in case

$$\begin{aligned}ag + c(1-g) &< agh + bg(1-h) + c(1-g) \\ ag &< agh + bg(1-h) \\ ag(1-h) &< bg(1-h) \\ a &< b\end{aligned}$$

which is guaranteed by assumption (1). Furthermore, we have $P(G|E) < / = / > P(G)$ just in case

$$\begin{aligned}ah + b(1-h) &< / = / > agh + bg(1-h) + c(1-g) \\ (b-a)(1-h) + a &< / = / > (b-c)g - (b-a)gh + c \\ (b-a)(1-h+gh) &< / = / > (b-c)g + (c-a) \\ (b-a)(1-h+gh) &< / = / > (b-c)g + ((b-a)-(b-c)) \\ (b-a)(-h+gh) &< / = / > (b-c)(g-1) \\ (b-c)(1-g) &< / = / > (b-a)h(1-g) \\ (b-c)/(b-a) &< / = / > h\end{aligned}$$

By assumption (1), we have $0 < (b-c)/(b-a) < 1$. In particular, if $c = (a+b)/2$, we get $(b-c)/(b-a) = 1/2$.