Euro Jnl Phil Sci (2016) 6:1–7
DOI 10.1007/s13194-015-0115-z

ORIGINAL PAPER IN PHILOSOPHY OF SCIENCE

Confirmation, increase in probability, and partial discrimination: A reply to Zalabardo

William Roche[1]

Received: 6 October 2014 / Accepted: 27 March 2015 / Published online: 26 April 2015

Abstract There is a plethora of confirmation measures in the literature. Zalabardo considers four such measures: PD (Probability-Difference), PR (Probability-Ratio), LD (Likelihood-Difference), and LR (Likelihood-Ratio). He argues for LR and against each of PD, PR, and LD. First, he argues that PR is the better of the two probability measures. Next, he argues that LR is the better of the two likelihood measures. Finally, he argues that LR is superior to PR. I set aside LD and focus on the trio of PD, PR, and LR. The question I address is whether Zalabardo succeeds in showing that LR is superior to each of PD and PR. I argue that the answer is negative. I also argue, though, that measures such as PD and PR, on one hand, and measures such as LR, on the other hand, are naturally understood as explications of distinct senses of confirmation.

Keywords Confirmation · Confirmation measures · Increase in probability · Likelihood-Ratio measure · Partial discrimination · Probability-Difference measure · Probability-Ratio measure · Zalabardo

✉ William Roche
w.roche@tcu.edu

1 Department of Philosophy, Texas Christian University, Fort Worth, TX, USA

1 Introduction

There is a plethora of confirmation measures in the literature.[1] Zalabardo (2009) considers four such measures[2]:

$$\begin{aligned}
\text{Probability Difference}: \quad & \mathrm{PD}(H,E) &= p\,(H|E) - p\,(H) \\
\text{Probability Ratio}: \quad & \mathrm{PR}(H,E) &= \frac{p(H|E)}{p(H)} \\
\text{Likelihood Difference}: \quad & \mathrm{LD}(H,E) &= p(E|H) - p(E|\neg H) \\
\text{Likelihood Ratio}: \quad & \mathrm{LR}(H,E) &= \frac{p(E|H)}{p(E|\neg H)}
\end{aligned}$$

He argues for LR and against each of PD, PR, and LD. First, he argues that PR is the better of the two probability measures. Next, he argues that LR is the better of the two likelihood measures. Finally, he argues that LR is superior to PR.

I want to set aside LD and focus on the trio of PD, PR, and LR. LD has received some support in the literature, but nothing like the support received by PD, PR, and LR (or ordinally equivalent measures).

The question I want to address is whether Zalabardo succeeds in showing that LR is superior to each of PD and PR. I aim to show that the answer is negative. I also aim to show, though, that measures such as PD and PR, on one hand, and measures such as LR, on the other hand, are naturally understood as explications of distinct senses of confirmation.

2 PD versus PR

Zalabardo appeals to a case given in Schlesinger (1995). Zalabardo writes:

> Schlesinger asks us to compare two scenarios. In the first, we consider a type of aircraft which is regarded as extremely safe, with a $1/10^9$ probability of crashing in a single flight. However, further inspection of the structure of the aircraft reveals a flaw as a result of which the probability of one of these planes crashing is actually 1/100. The second scenario concerns troops landing gliders behind enemy lines. We start from the assumption that someone taking part in one of these operations has a 26 % chance of perishing, but one day the commander announces that owing to peculiar weather conditions the risk has increased from 26 % to 27 %. (pp. 631–632)

[1] See Roche and Shogenji (2014) for a list of the main confirmation measures in the literature. See Roche (2014) for an expanded list.

[2] All references to Zalabardo are to Zalabardo (2009).

Zalabardo continues:

> [T]he degree to which the inspection of the aircraft confirms the hypothesis of a plane crash is intuitively much higher than the degree to which the unusual weather conditions confirm the hypothesis of a glider mission resulting in death. (p. 632)

Zalabardo then shows that this (claim about degree of confirmation) is true on PR but not on PD.

This argument has some force. But it is far from conclusive. This can be seen by considering a modified version of Schlesinger's case. There are two scenarios. Scenario 1: a plane P is set for takeoff; H is the proposition that P will crash shortly after takeoff; $p(H)=1/10^9$; some evidence E comes in to the effect that P has a certain structural flaw; $p(H \mid E)=1/100$. Scenario 2: a soldier S is set to embark on a mission in which gliders are to be landed behind enemy lines; H^* is the proposition that S will die during the mission; $p(H^*)=1/100$; some evidence E^* comes in to the effect that the weather conditions are extremely unsafe for flying; $p(H^* \mid E^*)=99/100$. Here, intuitively, the degree to which E confirms H is less than, in fact, much less than, the degree to which E^* confirms H^*. But $\mathrm{PR}(H, E)=10^7 >> 99=\mathrm{PR}(H^*, E^*)$ whereas $\mathrm{PD}(H, E)\approx 0.010 << 0.98=\mathrm{PD}(H^*, E^*)$.[3]

It might help to consider a case not involving potential plane crashes or potentially fatal military missions. Imagine a rather strange deck of cards: there are 10000 cards; 99 of the cards are diamond face cards; 1 of the cards is a heart face card; the remaining 9900 cards are black non-face cards. Suppose the deck is shuffled and a card is randomly drawn. Let E be the proposition that the card drawn is a face card, H be the proposition that the card drawn is a diamond, and H^* be the proposition that the card drawn is a heart. H's prior probability is very low: $p(H)=99/10000$. So too is H^*'s prior probability: $p(H^*)=1/10000$. But H's posterior probability is very high whereas H^*'s posterior probability is still very low: $p(H \mid E)=99/100$ while $p(H^* \mid E)=1/100$. It is not implausible, prima facie, that the degree to which E confirms H is greater than the degree to which E confirms H^*. This is borne out by PD: $\mathrm{PD}(H, E)\approx 0.980 >> 0.010 \approx \mathrm{PD}(H^*, E)$. But it is not borne out by PR: $\mathrm{PR}(H, E)=100=\mathrm{PR}(H^*, E)$.

Zalabardo, at any rate, is in no position to claim that PR's ordering in this case is the better of the two. For, on LR—Zalabardo's preferred measure—the degree to which E confirms H is greater than the degree to which E confirms H^*: $\mathrm{LR}(H, E)=9901 >> 101=\mathrm{LR}(H^*, E)$.

Neither the modified version of Schlesinger's case nor the card case is meant to tell decisively in favor of PD's superiority over PR. The point is just that Zalabardo's argument against PD, though not without force, is not the final word on the issue of PD versus PR.

PD and PR are naturally understood as measuring degree of confirmation in terms of degree of *increase in probability*.[4, 5] If the degree of increase in H's probability due to E

[3] LR, like PR, yields the (prima facie) implausible result that the degree to which E confirms H is greater than, in fact, much greater than, the degree to which E^* confirms H^*: $\mathrm{LR}(H, E)\approx 10101000 >> 9801=\mathrm{LR}(H^*, E^*)$.
[4] If Joyce is right, then this is true of any (Bayesian) confirmation measure. He writes: "All Bayesians agree that the degree to which [E] counts as evidence for … [H] [i.e., the degree to which E confirms H] for a given person is a matter of the extent to which learning [E] would increase … her confidence in [H]" (1999, p. 205).
[5] Similarly, PD and PR are naturally understood as measuring degree of *dis*confirmation in terms of degree of *decrease* in probability.

is large, then the degree of confirmation is large. If the degree of increase in H's probability due to E is middling, then the degree of confirmation is middling. And so on. PD measures degree of increase in H's probability by the *difference* between H's posterior and prior probabilities whereas PR measures degree of increase in H's probability by the *ratio* of H's posterior and prior probabilities. My worry with Zalabardo's argument (for PR's superiority over PD) can be put as follows: Schlesinger's case tells at least to some extent in favor of the ratio approach to measuring degree of increase in probability, but, at the same time, both the modified version of Schlesinger's case and the card case tell at least to some extent in favor of the difference approach, so it is far from clear, to say the least, that the ratio approach is the better of the two.

I turn now to Zalabardo's argument for LR's superiority over PR.

3 PR versus LR

Zalabardo's argument for LR's superiority over PR starts with a case. He writes:

> Consider the degree to which a diagnosis of asthma is supported by two standard symptoms: wheezing and a dry cough. Both symptoms have a very high ratio of true positives: most people with asthma wheeze and most people with asthma have a dry cough. Let's assume that the true-positive ratio is identical in each case. However, with respect to false positives, the two symptoms rate very differently. Very few people who don't have asthma wheeze, whereas quite a few people who don't have asthma have a dry cough. Hence wheezing and a dry cough have the same ratio of true positives, while wheezing has a significantly lower ratio of false positives than a dry cough does. (p. 633)

He continues:

> I want to suggest that a plausible theory of confirmation should yield the result that the features of the example that we have described suffice for concluding that wheezing confirms a diagnosis of asthma to a higher degree than a dry cough does. (p. 633)

He then generalizes and concludes that any adequate confirmation measure should meet the following condition:

(1) If $p(E \mid H)=p(E^* \mid H)$ and $p(E \mid \neg H)<p(E^* \mid \neg H)$, then E confirms H to a greater degree than E^* confirms H.

Note that this condition involves two pieces of evidence (E and E^*) but just one hypothesis (H).

LR meets condition (1). But, as Zalabardo notes, so too does PR. Hence condition (1) is neutral between LR and PR.

Now consider a similar but slightly different condition:

(2) If $p(E \mid H)=p(E^* \mid H^*)$ and $p(E \mid \neg H)<p(E^* \mid \neg H^*)$, then E confirms H to a greater degree than E^* confirms H^*.

This condition, like condition (1), involves two pieces of evidence (E and E^*). But condition (2), unlike condition (1), involves two hypotheses (H and H^*) as opposed to just one (H).

Zalabardo contends that, despite this difference, condition (2)—like condition (1)—should be met by any adequate confirmation measure. He writes:

> I want to argue next that intuition sanctions the same verdict on the importance of false positives when we are comparing the degree to which two pieces of evidence confirm different hypotheses. (p. 633)

> I want to suggest that intuition dictates that when we compare the degree to which wheezing supports a diagnosis of asthma with the degree to which weight loss supports a diagnosis of lung cancer, we should draw the same conclusion as when we compared wheezing and a dry cough as evidence for asthma. Wheezing has the same ratio of true positives with respect to asthma as weight loss does with respect to lung cancer, but the former has a lower false-positive ratio than the latter does. Hence a plausible account of confirmation should treat wheezing as confirming the asthma hypothesis to a higher degree than weight loss confirms the lung cancer hypothesis. (p. 634)

It turns out that LR meets condition (2) but PR does not. Zalabardo concludes that LR is superior to PR.

Is Zalabardo right that condition (2) should be met by any adequate confirmation measure? There is reason for answering in the negative. There are probability distributions on which each of (a)-(d) holds[6]:

(a) $p(E \mid H)=1=p(E^* \mid H^*)$
(b) $p(E \mid \neg H)=0.000005<0.00001=p(E^* \mid \neg H^*)$
(c) $p(H \mid E)\approx 1>0.99=p(H)$
(d) $p(H^* \mid E^*)\approx 1>0.050\approx p(H^*)$

Given (a) and (b), it follows by condition (2) that the degree to which E confirms H is greater than the degree to which E^* confirms H^*. It turns out, in fact, that on LR the degree to which E confirms H is *much* greater than the degree to which E^* confirms H^*:

$$\mathrm{LR}(H,E) = 200,000 >> 100,000 = \mathrm{LR}(H^*,E^*)$$

But, prima facie, it is not implausible that given (c) and (d), the degree to which E confirms H is less than, in fact, much less than, the degree to which E^* confirms H^*.

It is worth noting that:

$$\mathrm{PD}(H,E)\approx 0.010 << 0.950\approx \mathrm{PD}(H^*,E^*)$$
$$\mathrm{PR}(H,E)\approx 1.010 << 20.121\approx \mathrm{PR}(H^*,E^*)$$

So on each of PD and PR, unlike on LR, the degree to which E confirms H is less than—much less than—the degree to which E^* confirms H^*.

[6] For example: $p(E\wedge E^*\wedge H\wedge H^*)$ = 8/161; $p(E\wedge\neg E^*\wedge H\wedge\neg H^*)$=15139/16100; $p(E\wedge\neg E^*\wedge\neg H\wedge\neg H^*)$ = 1/20000000; $p(\neg E\wedge E^*\wedge\neg H\wedge\neg H^*)$ = 153/16100000; $p(\neg E\wedge\neg E^*\wedge\neg H\wedge\neg H^*)$ = 32169239/3220000000.

The lesson is clear: it is far from obvious, to say the least, that Zalabardo is right that condition (2) should be met by any adequate confirmation measure.

4 LR and confirmation as partial discrimination

It is a commonplace in Bayesian confirmation theory to distinguish between confirmation in the sense of *incremental* confirmation and confirmation in the sense of *absolute* confirmation. This is a commonplace for a reason: each sense of confirmation captures a significant respect in which E can be related to, evidentially, H. Perhaps there are additional distinctions to be made. Perhaps, in particular, there are distinctions to be made with respect to incremental confirmation. Perhaps there is incremental confirmation in the sense of increase in probability but also incremental confirmation in various other senses.[7] And perhaps LR is best understood not as an explication of incremental confirmation in the sense of increase in probability, but as an explication of incremental confirmation in some other sense. I want to close by suggesting that LR is perhaps best understood as an explication of confirmation as partial discrimination.[8]

Suppose you are building a test for disease D. Let H be the proposition that S has D (where S is some subject). Let E be the proposition that the test says that S has D. The best you can hope for is that E *fully* discriminates between H and $\neg H$ in that $p(E \mid H)=1$ while $p(E \mid \neg H)=0$.[9] If you ran the test and it turned out that E, then this would tell you definitively that H. If, instead, you ran the test and it turned out that $\neg E$, then this would tell you definitively that $\neg H$. Suppose, though, full discrimination is not in reach. But you have two rather attractive options: build a test such that $p(E \mid H)=1$ while $p(E \mid \neg H)=1/10$; build a test such that $p(E \mid H)=1$ while $p(E \mid \neg H)=1/20$. Each of the two tests would fall short of the ideal. But the second test would come closer and thus would be preferable (other things being equal). You thus build the second test.

Now suppose your friend builds a test for a different disease: D*. Let H^* be the proposition that S has D*. Let E^* be the proposition that the test says that S has D*. Suppose your friend's test is such that $p(E^* \mid H^*)=1$ while $p(E^* \mid \neg H^*)=1/10$. There is a clear sense in which you have the better test: E comes closer to full discrimination (with respect to H and $\neg H$) than does E^* (with respect to H^* and $\neg H^*$) so that the *degree* to which E *partially* discriminates between H and $\neg H$ is greater than the *degree* to which E^* *partially* discriminates between H^* and $\neg H^*$. This is true even if H's prior probability is significantly greater than H^*'s prior probability and, in part because of this, the degree of increase in H's probability due to E is less than the degree of increase in H^*'s probability due to E^*.

How does all this relate to LR? A natural way of measuring the degree to which E partially discriminates between H and $\neg H$ is by the ratio of $p(E \mid H)$ and $p(E \mid \neg H)$. LR is thus naturally understood as measuring confirmation in terms of partial discrimination. Recall condition (2). This condition, though implausible as an adequacy condition on

[7] Joyce (1999, Ch. 6, sec. 6.4) takes a pluralistic approach to confirmation (or evidential support). I am using the term "incremental" broadly so that "probative" confirmation (Hájek and Joyce 2008) would be a kind of incremental confirmation.

[8] I owe this idea to Roush (2005, Ch. 5). But I develop the idea somewhat differently than Roush does.

[9] $p(E \mid H)=1$ while $p(E \mid \neg H)=0$ iff $p(\neg E \mid \neg H)=1$ while $p(\neg E \mid H)=0$. So E fully discriminates between H and $\neg H$ iff $\neg E$ fully discriminates between $\neg H$ and H.

measures of confirmation as increase in probability, makes perfect sense as an adequacy condition on measures of confirmation as partial discrimination. If $p(E \mid H)=p(E^* \mid H^*)$ while $p(E \mid \neg H)<p(E^* \mid \neg H^*)$, then E comes closer to the ideal of full discrimination (with respect to H and $\neg H$) than does E^* (with respect to H^* and $\neg H^*$) and so the degree to which E confirms H is greater than the degree to which E^* confirms H^*.[10]

I see no need to try to force a choice between confirmation in the sense of increase in probability and confirmation in the sense of partial discrimination. Each sense, it seems, captures a significant respect in which E can be related to, evidentially, H.

Acknowledgments I wish to thank an anonymous reviewer for a helpful suggestion and Tomoji Shogenji for extremely helpful comments on prior versions of the paper.

References

Hájek, A., & Joyce, J. (2008). Confirmation. In S. Psillos & M. Curd (Eds.), *The Routledge companion to philosophy of science* (pp. 115–128). London: Routledge.
Joyce, J. (1999). *The foundations of causal decision theory*. Cambridge: Cambridge University Press.
Roche, W. (2014). A note on confirmation and Matthew properties. *Logic & Philosophy of Science, XII*, 91–101.
Roche, W., & Shogenji, T. (2014). Dwindling confirmation. *Philosophy of Science, 81*, 114–137.
Roush, S. (2005). *Tracking truth: Knowledge, evidence, and science*. Oxford: Oxford University Press.
Schlesinger, G. (1995). Measuring degrees of confirmation. *Analysis, 55*, 208–212.
Zalabardo, J. (2009). An argument for the likelihood-ratio measure of confirmation. *Analysis, 69*, 630–635.

[10] If LR is to be understood as measuring degree of confirmation in terms of partial discrimination, then, presumably, the same is true of LD. It might be that standard objections to LD need to be rethought.