Affirming the Explanandum

Borut Trpin

Forthcoming in *Analysis*

Abstract

Affirming the consequent is an inferential pattern, in which one infers the antecedent of a given conditional from its consequent. Abductive inference is structurally similar: Given some evidence, one infers a hypothesis that explains the evidence. I show that a Bayesian analysis of affirming the consequent helps us understand under which conditions abduction may be justified. This provides a Bayesian vindication of explanatory inference.

Keywords: Affirming the Consequent, Argumentation, Abductive Inference, Bayesian Reasoning, Explanatory Considerations

Affirming the Explanandum

Borut Trpin

1. Introduction

Affirming the consequent (AC) denotes an inferential pattern in which one infers the antecedent A from the consequent C and the conditional 'If A, then C'. In logical notation (with $\supset$ representing material implication):

$$\begin{array}{ll} & C \\ & A \supset C \\ \hline \therefore & A \end{array}$$

Although AC is deductively invalid (the premisses do not necessarily lead to the conclusion), it is easy to find cases where it seems that the conclusion is plausible because of (and not only despite) the premisses. For instance, the conclusion that it rains seems rather plausible given that the streets are wet and that the streets are wet if it rains. Of course, it may have just stopped raining or there could be a leaking fire hydrant, but because wet streets and rain largely overlap, the conclusion appears plausible. Besides these general intuitions, which may or may not be shared, it is also possible to find plenty of psychological evidence that people often accept the conclusion of AC (e.g. Thompson 1994, Oaksford et al. 2000).

It is one thing to point out that some inferential pattern is common in a descriptive sense. But could AC ever be acceptable from a normative viewpoint? To answer this question, it is helpful to restate the problem in probabilistic terms and apply the following Bayesian analysis (following Eva et al. 2018): Suppose you learn C and $A \rightarrow C$, where $\rightarrow$ represents the indicative conditional connective. Do you need to also become more confident of A in order to remain probabilistically coherent? If yes, then AC is reasonable. Else, it is not.

Next, let us focus on abductive inference by considering an example:

> You happen to know that Tim and Harry have recently had a terrible row that ended their friendship. Now someone tells you that she just saw Tim and Harry jogging together. The best explanation for this that you can think of is that they made up. You conclude that they are friends again. (Douven 2021)

In this case, one infers (or in probabilistic terms, becomes more confident of) a hypothesis H: 'Tim and Harry are friends again' because one learned evidence E: 'Tim and Harry are

jogging together' and because H provides an explanation of E. And indeed, the conclusion that Tim and Harry are friends seems reasonable, even though it is deductively invalid (they could be jogging together for some other reason). Just like before, it is also easy to find cases where an abductive inference is misguided. For instance, suppose that a student yawns during a lecture. It would be wrong to conclude that the student is sleepy if they are merely bored (e.g., because the lecture is on Hegel's phenomenology, which many students find hard to engage with). As with AC, it seems that restating the problem in probabilistic terms might also help us understand when abductive inference is reasonable from a normative viewpoint and when it is not. In fact, doing so might be quite straight-forward because abductive inference is structurally similar to AC (Pfister 2022). This becomes clear when Peirce's well-known characterization of abduction is considered:

> The surprising fact, C, is observed.
>
> But if A were true, C would be a matter of course.
>
> Hence, there is a reason to suspect that A is true. (Peirce 1935: 5.189)

The obvious differences in comparison to AC are that we operate with an explanans (i.e., that which does the explaining) instead of A and with an explanandum (i.e., that which is to be explained) instead of C, and that the conditional, which represents the explanation of C by A, is subjunctive. However, as I will argue, these differences are minor when this specific form of abductive inference is analysed in a Bayesian framework.

Because AC can be normatively reasonable and because at least some abductive inferences may be characterized as analogous to AC, it follows that this type of abductive inference can also be normatively reasonable. The approach of Eva and Hartmann (2018) turns out to be very helpful in this context: it affords an analogous investigation of the conditions under which it is rational for a Bayesian agent to infer an explanans from an explanandum. I will call this inferential pattern *affirming the explanandum* (AE) in analogy to the well-known AC. I will explore under which conditions AE is reasonable and whether and to what degree these results vindicate abduction in a Bayesian context. In other words, I will outline in which cases this type of abductive inference is part of Bayesian inference.

In the next section (Section 2), the focus will be on the cases where the explanandum is learned with certainty. I will then move on to cases where the explanandum E is uncertain (Section 3) to highlight that confirmation does not necessarily warrant abduction. To wrap the discussion up, I will briefly discuss how my analysis vindicates abductive inference in a Bayesian context (Section 4).

2. Learning the explanandum with full certainty

Let us start the investigation with a simple case. Following the approach of Eva and Hartmann (2018), I introduce two binary propositional variables H and E (in italic script). The variable H represents the hypothesis that does the explaining (the explanans), and E represents the evidence that is to be explained (the explanandum). The variables have the values H and ¬H, and E and ¬E (in roman script), which respectively correspond to whether the explanans (H) or the explanandum (E) hold. An agent's prior probability distribution P (that is, prior to learning about the explanandum) can then be represented with the parameters:

$$P(\mathrm{H}) = h$$

$$P(\mathrm{E}|\mathrm{H}) = p, \quad P(\mathrm{E}|\neg\mathrm{H}) = q. \tag{1}$$

The prior distribution P over the variables H and E is then given by

$$\begin{aligned} P(\mathrm{H\&E}) &= hp, \quad & P(\mathrm{H\&\neg E}) &= h\bar{p}, \\ P(\neg\mathrm{H\&E}) &= \bar{h}q, \quad & P(\neg\mathrm{H\&\neg E}) &= \neg h\bar{p}, \end{aligned} \tag{2}$$

where $\bar{x}$ is used as a shorthand for $1 - x$. Throughout the paper, I also assume that all relevant parameters of the prior distribution are non-extreme: $0 < h < 1$, $0 < p < 1$, $0 < q < 1$. Returning to the example by Douven (2021) mentioned above, P represents the state of an agent before learning that Tim and Harry are jogging together (E). The agent thinks it is unlikely that they are friends ($P(\mathrm{H})$ is low) and that $p > q$, regardless of their specific values (it is more likely that they jog together if they are friends than if they are not).

Next, the agent learns E: Tim and Harry are jogging together, and reasons that E is best explained by H ('If H were the case, E would be a matter of course.'). This prompts them to update P to a posterior probability distribution P', which may be represented by the corresponding primed variables h', p', q':

$$\begin{aligned} P'(\mathrm{H\&E}) &= h'p', \quad & P'(\mathrm{H\&\neg E}) &= h'\bar{p}', \\ P'(\neg\mathrm{H\&E}) &= \bar{h}'q', \quad & P'(\neg\mathrm{H\&\neg E}) &= \neg h'\bar{p}'. \end{aligned} \tag{3}$$

Note that P' is constrained by the information the agent learned (E) and explanatorily reasoned about (that H explains E). The simplest way to include these constraints is therefore to set $P'(\mathrm{E}) = h'p' + \bar{h}'q' = 1$ (constraint AE1), and $P'(\mathrm{E}|\mathrm{H}) = p' = 1$ (constraint AE2). Here, AE1 simply amounts to learning the explanandum, while AE2 represents a probabilistic interpretation of considering E being a matter of course if H

were true. It is then possible to show the following (proof omitted as I am reusing a result by Eva and Hartmann 2018: 810, Proposition 3):

Proposition 1. An agent considers the propositions H and E and has a prior probability distribution P according to Eq. 2 defined over them. Learning the explanandum E by setting $P'(\mathrm{E}) = 1$ and reasoning that E would be a matter of course in case H were true by setting $P'(\mathrm{E}|\mathrm{H}) = 1$ and minimizing the Kullback-Leibler divergence between P' and P, then implies that the new probability of the explanans H, that is $P'(\mathrm{H})$, equals $P(\mathrm{H}|\mathrm{E})$.

This is an interesting result: in the simple case where an agent learns E with certainty and reasons that H explains E by making it completely expected, abduction turns out to be reasonable exactly when the posterior probability of H after learning E is high. The same result also follows if one interprets AE1 and AE2 as learning E and a material conditional $\mathrm{H} \supset \mathrm{E}$ and performs standard Bayesian conditionalization on their conjunction. This is because $\mathrm{E}\&(\mathrm{H} \supset \mathrm{E}) \Leftrightarrow \mathrm{E}$. In our running example, it is reasonable to infer that Tim and Harry are friends because this explains their jogging together exactly when their jogging together confirms their renewed friendship. In these simple cases abductive reasoning turns out to be equivalent to simple Bayesian conditionalisation on the explanandum. Abductive inference can therefore be completely warranted, but only if it does not diverge from standard Bayesian updating on the evidence that needs to be explained. The explanatoriness itself, however, turns out to be evidentially irrelevant (concurring with Roche and Sober 2013).

3. Uncertain explanandum

Nevertheless, this does not mean that the prospects for Bayesian explanationism are doomed. For one, I have so far only considered abductive inference when an agent becomes fully certain of the explanandum. However, as Lindley (1985: 104) warns, it is a good idea to leave a little room for doubt and not assign extreme probabilities to uncertain events (i.e., to anything but logical tautologies and contradictions).

Let us therefore see what happens in case the explanandum is not learned with full certainty. Douven's (2021) introductory example again turns out to be helpful. Suppose that an agent learns that Tim and Harry were jogging together via testimony (as in Douven's example), but does not completely trust the report. Or perhaps the agent saw Tim and Harry jogging together but from rather far and in poor lighting conditions. In any case, the agent becomes more but not fully certain of the surprising explanandum. The agent then reasons that if Tim and Harry are friends again, then their jogging would be a matter of course. The learning constraints in such scenarios are $P(\mathrm{E}) < P'(\mathrm{E}) < 1$ (constraint AE3) and $P'(\mathrm{E}|\mathrm{H}) = 1$ (constraint AE4). When is it then reasonable to conclude that they buried the hatchet? More generally, when is it reasonable to infer

the explanans due to its being a matter of course given an uncertain explanandum? The result depends on the prior probability distribution and on how certain the agent becomes of the explanandum E (proof in Appendix):

Proposition 2. An agent considers the propositions H and E and has a prior probability distribution P according to Eq. 2 defined over them. Learning the explanandum E by setting $P'(\mathrm{E}) > P(\mathrm{E})$ and reasoning that E would be a matter of course in case H were true by setting $P'(\mathrm{E}|\mathrm{H}) = 1$ and minimizing the Kullback-Leibler divergence between P' and P implies that the new probability of the explanans H increases iff $P'(\mathrm{E}) > P(\mathrm{E})/P(\mathrm{E}|\mathrm{H})$.

Note that the same result could also be obtained by Jeffrey conditioning on E and the material conditional $\mathrm{H} \supset \mathrm{E}$ (see Proposition 9 in Eva and Hartmann 2018). In a simplified sense, the result shows that abductive inference is only reasonable if the agent becomes more certain of the explanandum E than what they initially thought the relevance of the explanans H for the explanandum E was. If the agent initially thought that H more or less entails E (i.e., if $P(\mathrm{E}|\mathrm{H}) = 1 - \epsilon$), then abductive inference to H from E is warranted as soon as the agent becomes more certain of E than they initially were. The less the agent initially thought H was relevant for E, however, the more certain they need to become of E for the inference to be warranted. Moreover, it is also easy to see that it is never reasonable to become more certain of the explanans H if E and H were initially assumed to be negatively correlated (i.e., if $P(\mathrm{E}|\mathrm{H}) < P(\mathrm{E}|\neg\mathrm{H})$). In this case the threshold that warrants increased confidence in the explanans H is $P(\mathrm{E})/P(\mathrm{E}|\mathrm{H}) > 1$ and it is impossible to become more certain of E than that. Affirming the explanandum thus turns out to be predestined to fail in such cases.

To make the result easier to understand, suppose the agent initially thought Tim and Harry were somewhat unlikely to go jogging together, e.g., $P(\mathrm{E}) = .4$. The agent then learns some evidence which suggests that they are likely jogging together and sets $P'(\mathrm{E}) = .7$ and, moreover, reasons that if they are friends again, then their jogging would be a matter of course: $P'(\mathrm{E}|\mathrm{H}) = 1$. Proposition 2 tells that affirming the explanandum is in this case only reasonable if the agent initially thought $P(\mathrm{E}|\mathrm{H}) > 4/7 \approx .57$, that is, if the agent initially thought that jogging was more likely if they reestablished their friendship. Importantly, if the agent initially thought that the reestablished friendship would increase the chances of their jogging but not by much, e.g., $P(\mathrm{E}|\mathrm{H}) = .55 > P(\mathrm{E}) = .4$, then the inference to the friendship is not warranted because the posterior probability of E is not high enough. This means that it might be rational to (i) find it likely that Tim and Harry are jogging, (ii) to initially think that jogging is more likely if Tim and Harry are friends, (iii) to come to think that their running would be completely expected if they are friends, and yet despite (i)-(iii), it may still not be warranted to become more confident that they are friends again. Even though the agent

may come to think that their friendship perfectly accounts for the (uncertain) jogging, this conclusion was already doomed from the start. This also highlights that although H and E may be positively correlated and therefore in a mutually confirmatory relation, an abductive inference from E to H will still not always be warranted. In other words, even if E confirms H and H explains E, it is not always rational to infer the explanans H – the standards of reasonable abductive inference are stricter than the standards of confirmatory relations. And in some cases (i.e., given specific priors), abduction is not an option at all.

To sum up: this brief investigation shows that in case the explanandum is not learned with full certainty, Bayesians infer the explanans only when the explanation of E by H was already assumed to be plausible enough and when E has been learned with enough certainty. This makes sense: if one is not certain enough of Tim and Harry jogging or if they initially thought that their jogging does not make enough sense in light of the renewed friendship, then it makes no sense to infer that they are friends again even if this would provide a perfect explanation of the uncertain observation. A Bayesian characterization of this type of abduction may therefore be given in the following way:

> The surprising fact, E, is observed, which prompts an increase in its probability.
>
> But if H were true, E would be a matter of course.
>
> Hence, if E has become probable enough and H was also initially already assumed to be relevant enough for E, there is a reason to suspect that H is true.

4. Conclusion

My investigation shows that abduction is conditionally compatible with Bayesianism: abductive inference is compatible with Bayesianism whenever learning the explanandum and inferring the conditional dependence of the explanandum on the explanans leads to a rational increase in the probability of the explanans. Peirce's characterization of abduction helps us understand the type of inference that is ubiquitous in everyday and scientific reasoning (Lombrozo 2006), while a Bayesian analysis helps us understand why and when it is reasonable. My investigation also suggests why many (e.g. Cabrera 2017, Climenhaga 2017, van Fraassen 1989, Roche and Sober 2013) consider abduction to be distinct from Bayesian inference – because just the fact that H explains E does not always suffice to become more certain of E. Yet it often does and my analysis provides the conditions under which it does for at least one type of abductive reasoning (AE).

My focus has been on abduction, but this also helps us understand the role of inference to the best explanation (IBE) in Bayesianism. If there are multiple potential explanantia

H_i for a single explanandum E (i.e., several high conditional probabilities $P(E|H_i)$), then IBE may be characterized as a conjunction of AE for each H_i with respect to the same E and the fact that some H_i become(s) most plausible – depending on the prior probabilities and the degree to which the agent becomes certain of the explanandum. IBE may in this sense then also be seen as a vital component of the standard Bayesian toolkit.

Note also that my aim is to show that abduction is compatible with standard Bayesianism, which is also why I do not address cases where one does not become sure whether E would be a matter of course if H were true (i.e., when $P'(E|H) < 1$). In these cases, the conditional relation between H and E is uncertain and as such demands a non-standard Bayesian treatment of conditional learning on which there is no clear consensus yet (for three varying views, see e.g. Douven 2012, Eva et al. 2020, Günther and Trpin 2023). This need not be seen as a limitation because Peirce's guiding characterization of abduction also operates with the assumption that E would be 'a matter of course' if H were true. In future work it would be nevertheless interesting to see what follows when this assumption is relaxed.

It is worth noting that this is in contrast to well-known critics of Bayesian explanationism such as van Fraassen (1989) and Roche and Sober (2013) (for an overview of the discussion and a more general defense of explanationism, see Lange 2023). The approach also differs from Weisberg's (2009) where explanatory considerations are taken into account in informing the priors. Here, I have instead shown that under the appropriate conditions abductive inference is simply part of Bayesian inference as such. Finally, my approach shows that one may take explanatory considerations into account in a probabilistic framework without necessarily deferring to non-Bayesian forms of belief updating, such as those considered by Douven (2013) and Trpin and Pellert (2019). Of course, there may also be other reasonable ways of spelling out abduction, which I did not address. However, my investigation shows that at least in some senses (that is, at least in the sense of AE) abduction is part of Bayesianism. Paraphrasing Climenhaga (2017), inference to the best explanation is made coherent insofar it corresponds to Bayesian inference, which it very well may.[1]

LMU Munich
Germany
borut.trpin@lrz.uni-muenchen.de

Funding

This work was supported by the Arts and Humanities Research Council and The Deutsche Forschungsgemeinschaft [grant numbers HA 3000/20-1, HA 3000/21-1].

[1] Thanks to two anonymous referees for helpful comments.

Appendix

With the prior and posterior probability distributions as defined in Eqs. 2 and 3, and reusing the result by Eva and Hartmann 2018: Supplementary Material, Eq. 27:

$$P'(\mathrm{H}) = h' = P(\mathrm{H}|\mathrm{E})P'(\mathrm{E}) = \frac{hp}{e}e', \tag{4}$$

where $P(\mathrm{E}) := hp + \bar{h}q$ and $P'(\mathrm{E}) = h'p' + \bar{h}'q'$. It is then straightforward to see that

$$\Delta_H := P'(\mathrm{H}) - P(\mathrm{H}) = h' - h = \frac{hp}{e}e' - h > 0 \tag{5}$$

$$\Leftrightarrow e' > \frac{e}{p} \tag{6}$$

□

References

Cabrera, F. 2017. Can there be a Bayesian explanationism? On the prospects of a productive partnership. *Synthese* 194(4): 1245—72.

Climenhaga, N. 2017. Inference to the best explanation made incoherent. *Journal of Philosophy* 114(5): 251—73.

Douven, I. 2012. Learning conditional information. *Mind & Language* 27(3): 239-–63.

Douven, I. 2013. Inference to the best explanation, Dutch books, and inaccuracy minimisation. *Philosophical Quarterly* 63(252): 428–44.

Douven, I. 2021. Abduction. In *The Stanford Encyclopedia of Philosophy* (Summer 2021 edn.), ed. E.N. Zalta. <https://plato.stanford.edu/archives/sum2021/entries/abduction/>.

Eva, B. and S. Hartmann. 2018. Bayesian argumentation and the value of logical validity. *Psychological Review* 125(5): 806–21.

Eva, B., S. Hartmann and S.R. Rad. 2020. Learning from conditionals. *Mind* 129(514): 461—508.

Günther, M. and B. Trpin. 2023. Bayesians still don't learn from conditionals. *Acta Analytica* 38: 439—51.

Lange, M. 2023. A false dichotomy in denying explanatoriness any role in confirmation. Noûs.

Lindley, D.V. 1985. *Making Decisions* (Second ed.). London, New York, Brisbane, Toronto, Singapore: John Wiley & Sons.

Lombrozo, T. 2006. The structure and function of explanations. *Trends in Cognitive Sciences* 10(10): 464-–70.

Oaksford, M., N. Chater and J. Larkin. 2000. Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26(4): 883—99.

Peirce, C.S. 1931-1935. *The Collected Papers of Charles Sanders Peirce, Volume 1-6*. Cambridge, MA: Harvard University Press.

Pfister, R. 2022. Towards a theory of abduction based on conditionals. *Synthese* 200(3): 1–30.

Roche, W. and E. Sober. 2013. Explanatoriness is evidentially irrelevant, or inference to the best explanation meets Bayesian confirmation theory. *Analysis* 73(4): 659—68.

Thompson, V.A. 1994. Interpretational factors in conditional reasoning. *Memory & Cognition* 22(6): 742-–58.

Trpin, B. and M. Pellert 2019. Inference to the best explanation in uncertain evidential situations. *British Journal for the Philosophy of Science* 70(4): 977-–1001.

van Fraassen, B. 1989. *Laws and Symmetry*. Oxford: Oxford University Press.

Weisberg, J. 2009. Locating IBE in the Bayesian framework. *Synthese* 167(1): 125—43.