DEFEAT RECONSIDERED

JAKE CHANDLER*

Abstract It appears to have gone unnoticed in the literature that Pollock's widely endorsed analysis of evidential defeat entails a remarkably strong symmetry principle, according to which, for any three propositions D, E and H, if both E and D provide a reason to believe H, then D is a defeater for E's support for H if and only if, in turn, E is a defeater for D's support for H. After illustrating the counterintuitiveness of this constraint, a simple, more suitable, alternative to the Pollockian account is offered.

Keywords defeater - evidence - Pollock

The reasons that wed us to our worldviews are typically, in philosophical parlance, "defeasible": the inferences that we draw are often to be reneged on as new considerations come to light. To take a pair of stock examples, inspecting the first ten thousand widgets on a production line and finding these to be faultless (E_W) might provide reasonable grounds to hold that the remaining thousand are similarly defect-free (H_W). Similarly, upon observing an art installation through a peep hole, one might reasonably conclude from the fact that the artefacts on display appear red (E_A) that they are indeed red (H_A). But these conjectures would have to subsequently be retracted if one were to, respectively, discover of a further, previously uninspected, widget that it is defective (D_W) or find out that the installation is in fact illuminated by a red light (D_A).

*Address for correspondence: Center for Logic and Analytic Philosophy, HIW, KU Leuven, Kardinaal Mercierplein 2, 3000 Leuven, Belgium. Email: jacob.chandler@hiw.kuleuven.be

It is said that the last two propositions are "defeaters" of the relevant items of evidence with respect to their associated inferences.[1]

In his influential 'Defeasible Reasoning', the late John Pollock famously offered the following analysis of the familiar general phenomenon instantiated in the previous cases:

> Defeater: Where D and E are jointly consistent propositions, D is a defeater for E's support for H if and only if (i) E is a reason to believe H but (ii) $E \& D$ is not a reason to believe H.[2]

Thus, according to Pollock, D_W defeats the inference from E_W to the conclusion H_W because: (i) E_W provides a reason to believe H_W and (ii) $E_W \& D_W$ does *not* provide such a reason, since indeed the latter arguably provides a reason to believe $\neg H_W$. Similarly, D_A is alleged to defeat the inference from E_A to H_A because: (i) E_A provides a reason to believe H_A and (ii) $E_A \& D_A$ does *not* provide such a reason.

It is hard to overstate the influence that this analysis has had in the literature. Indeed, its correctness has been taken for granted in pretty much every discussion of the topic, both in philosophy and elsewhere, notably in artificial intelligence, where Pollock's work has left a lasting legacy. It may therefore come as somewhat of a surprise to find out that it faces a fairly straightforward kind of counterexample.

Indeed, consider: Outside the door to Sam's flat is a switch for the light on the landing of the floor below. Flipping the switch (E_S) typically causes the light to go on (H_S): E_S is therefore a prima facie reason to believe H_S. Of course, in the event of a power cut (D_S), E_S loses this probative force since, under such circumstances, whether or not the light is on is causally independent of the position of the switch. D_S, we would intuitively like to say, is a defeater for E_S with respect to H_S.

[1] According to standard terminology, D_W and D_A are "rebutting" and "undercutting" defeaters, respectively. The distinction, which will not preoccupy us here, hinges on the fact that D_W is also a reason to believe $\neg E_W$, whilst no such relation obtains between D_A and $\neg E_A$.

[2] See Pollock (1987, p. 284). This account is repeated, in a very slightly different form, in the popular epistemology textbook that he later co-authored with Joseph Cruz (Pollock & Cruz 1999, p. 195). This proposal finds its roots in the works of Chisholm; see Chisholm (1989, p. 53).

Now, to comply with local regulations, Sam's landlord has installed a backup power system that is activated in the event that the main system fails, automatically powering the light in the staircase to prevent tenants and their visitors from stumbling in the dark. So, just like E_S, D_S provides a reason to believe H_S. But it is worth noting an important asymmetry here: whilst D_S is a defeater for E_S with respect to H_S, E_S is clearly *not*, from a pre-theoretical viewpoint, a defeater for D_S with respect to the same proposition: the evidential connection that obtains between the occurrence of a power cut and the staircase light's being on is entirely unaffected by the position of the switch.

It should now be quite clear that Defeater leaves us in quite an unpleasant predicament. Indeed, either (i) $E_S \& D_S$ is a reason to believe H_S, or (ii) it is not. But we do not need to adjudicate here: either option spells trouble. If we assume (i), it follows by Defeater that D_S is not a defeater for E_S with respect to H_S, contrary to our intuitions. If we assume (ii), it follows this time, from the same principle, that E_S *is* a defeater for D_S with respect to H_S, again contrary to our intuitions. The analysis simply cannot be made square with our untutored judgments on this case.

This kind of scenario makes vivid the implausibility of an immediate consequence of Pollock's account, namely that given three propositions D, E and H, if both E and D provide a reason to believe H, then D is a defeater for E's support for H if and only if, in turn, E is a defeater for D's support for H.

So the question then immediately arises: Could one perhaps provide an alternative to Defeater that both handles the original cases and delivers the intuitive verdict concerning the light switch example? One suggestion that would not be of any help here would be to recast Defeater with a different negational scope, requiring, not that $E \& D$ not be a reason to believe H, but that *$E \& D$ be a reason not to believe H*. Indeed, in the switch case, $E \& D$ is clearly *not* a reason not to believe H. The upshot of the modified principle would be that, contrary to intuitions, D is not a defeater for E with respect to H. But there is one strikingly simple proposal that does fit the bill:

Defeater*: Where D and E are jointly consistent propositions, D

is a defeater for E's support for H if and only if D is a reason to not believe that E is a reason to believe H.

It is easy to see that DEFEATER* does not share its predecessor's misfortunes: the inference to the problematic principle no longer goes through. It also yields the correct epistemic story with respect to the light switch case. D_S provides grounds to hold that E_S is no reason to believe H_S. However E_S does *not* provide grounds to hold that D_S is no reason to believe H_S. We also recover the desired verdict with respect to the widget and art gallery cases: D_W (respectively D_A) provides grounds to hold that E_W (respectively E_A) is no reason to believe H_W (respectively H_A).

Acknowledgments

I am grateful to Luc Bovens, Igor Douven and Adam Rieger for helpful feedback on an earlier version of this paper.

References

Chisholm, R. (1989). *Theory of Knowledge, 3rd edition.* Englewood Cliffs: Prentice-Hall.

Pollock, J.L. (1987). Defeasible Reasoning. *Cognitive Science* 11, pp. 481–518.

Pollock, J.L. & J. Cruz (1999). *Contemporary Theories of Knowledge. 2nd Edition.* Oxford: Rowman & Littlefield.