

To establish CNESS-causation requires having a look at the counterfactual setting  $(M_{A_1 \leftarrow 0}, \bar{u})$ . In this setting we get that  $A_1 = 0$ ,  $A_2 = 1$ ,  $BH_1 = 0$ , and thus  $BH_2 = 1$  (as well as  $V = 1$ ). (Informally: if *Assassin<sub>1</sub>* had not shot, then *Assassin<sub>2</sub>*'s bullet would have hit and killed Victim.) Here  $A_1 = 0$  directly NESS-causes  $BH_1 = 0$ ,  $BH_1 = 0$  directly NESS-causes  $BH_2 = 1$  (since it forms a sufficient set together with  $A_2 = 1$  and  $A_2 = 1$  does not suffice on its own), and  $BH_2 = 1$  directly NESS-causes  $V = 1$ . Therefore  $A_1 = 0$  NESS-causes  $V = 1$  along  $p^* = \{A_1, BH_1, BH_2, V\}$ . (Take note of this surprising finding. We come back to it in Example 3.) Since  $p^* \not\subseteq p$ , we get that  $A_1 = 1$  CNESS-causes  $V = 1$  (whereas  $A_1 = 0$  does not CNESS-cause  $V = 1$  in the counterfactual setting, since  $p \subseteq p^*$ ).

To see that  $A_1 = 1$  HP-causes  $V = 1$ , it suffices to note that  $(M, \bar{u}) \models BH_2 = 0$  and  $(M, \bar{u}) \models [A_1 \leftarrow 0, BH_2 \leftarrow 0]V = 0$ . Lastly, I leave it to the reader to verify that  $A_2 = 1$  is not an HP-cause of  $V = 1$ , and nor is it a direct NESS-cause of anything. Because of the latter,  $A_2 = 1$  is not a NESS-cause or a CNESS-cause of anything either.

Modifying the BvH definition so that it uses NESS-causation instead of direct NESS-causation is not a solution, for the NESS definition itself is problematic, as the following example shows. (In the appendix I discuss one more example, a so-called ‘Frankfurt-case’, to show that BvH’s reliance on strategies as opposed to events forms another source of problems.)

**Example 3.** *We revisit the counterfactual setting of Example 2 in which *Assassin<sub>1</sub>* does not shoot, so that Victim is killed by *Assassin<sub>2</sub>*'s shot.*

We already established for this scenario that  $A_1 = 0$  NESS-causes  $V = 1$ . Thus if we use the NESS definition, we get the absurd result that *Assassin<sub>1</sub>* failing to shoot causes Victim to die. If we then supplement the example so that also BvH’s **Epistemic Condition** is fulfilled, we get that *Assassin<sub>1</sub>* comes out as being responsible for Victim’s death. (Imagine, for instance, that they mistakenly believe to be holding a flare gun that could sound a warning shot so that Victim ducks for cover to avoid *Assassin<sub>2</sub>*'s bullet.) We already established that  $A_1 = 0$  does not CNESS-cause  $V = 1$ , the reader may verify that the same holds for the HP-definition.

This leaves CNESS-causation and HP-causation as candidates for the **Causal Condition**. I use Halpern & Pearl’s own example to argue against HP-causation [15].

**Example 4 (Loader).** *“Suppose that a prisoner dies either if A loads B’s gun and B shoots, or if C loads and shoots his gun. Taking D to represent the prisoner’s death and making the obvious assumptions about the meaning of the variables, we have that  $D = 1$  iff  $(A = 1 \wedge B = 1) \vee C = 1$ . Suppose that in the actual context  $\bar{u}$ , A loads B’s gun, B does not shoot, but C does load and shoot his gun, so that the prisoner dies. Clearly  $C = 1$  is a cause of  $D = 1$ . We would not want to say that  $A = 1$  is a cause of  $D = 1$ , given that B did not shoot (i.e., given that  $B = 0$ ).” [emphasis added]*

I agree with Halpern and Pearl. A fortiori, A is not responsible for the prisoner’s death, even if A only loaded the gun because he was convinced that B would shoot. Now consider the following variant. In the original example, C’s shot is determined directly by the context. Imagine we add a little twist, so that C would only fire his gun if B did not, i.e., the equation for C is  $C = \neg B$ . The above reasoning regarding A still applies, and therefore I believe it is a mistake to all of a sudden consider  $A = 1$  a cause of  $D = 1$ . Yet  $A = 1$  now is an HP-cause of  $D = 1$  (as it appears in the HP-cause  $A = 1 \wedge B = 0$ ), and thus A would be considered responsible for the prisoner’s death. The CNESS definition avoids this result (as does the NESS definition): the only candidate sufficient set for  $D = 1$  of which  $A = 1$  could be a necessary part, is  $\{A = 1, B = 1\}$ . So the mere fact that  $B = 0$  in both versions of the example implies that  $A = 1$  is not a NESS cause of  $D = 1$  in either.

I leave a second counterexample to the HP definition for the appendix and refer the reader to [1] for a detailed critical examination of the HP definition. The alternative definition I there presented is in fact very similar to my CNESS definition, although the precise relation is the subject of further investigation.<sup>5</sup> This leads me to suggest adopting the CNESS definition for the **Causal Condition**.

## 5 The Epistemic Condition

Recall that the difference between HK and BvH’s **Epistemic Conditions** lies in whether an action minimizes the probability of the *outcome occurring* (HK) or of *it causing the outcome* (BvH). Given

<sup>5</sup>I tentatively conjecture that the CNESS definition implies my other definition, and not vice versa.