- **(Epistemic Condition)** *There exists* $a' \in \mathcal{R}(A)$ *so that* $\Pr(O = o|[A \leftarrow a]) > \Pr(O = o|[A \leftarrow a'])$.

In addition to disagreeing about the definition of causation, the HK definition also disagrees with the BvH definition about the epistemic condition: rather than requiring that the agent failed to minimize the probability of causing the outcome, the HK definition focuses on the agent failing to minimize the probability of the outcome simpliciter.

Note that both HK and BvH's epistemic condition satisfy our **Responsibility Schema**: an agent who believes that they failed to minimize a probability that they could have minimized, thereby also believes that they could have avoided satisfying the respective epistemic condition. Given that the epistemic condition is a necessary condition for being responsible, they also believe that they could have avoided being responsible for the actual outcome.

Let us apply the HK definition to Example 1. We already established that each $A_i = 1$ is an HP-cause of $V = 1$, so the **Causal Condition** is met. Further, as long as each assassin attributes a strictly positive probability that the other assassin may fail to shoot, we get that $\Pr(V = 1|[A_i \leftarrow 1]) > \Pr(V = 1|[A_i \leftarrow 0])$, so that the **Epistemic Condition** is satisfied as well. (What if the assassins are certain the other assassin will shoot? We come back to this in Section 5.) Therefore the HK definition also arrives at the correct verdict for this example.

4 The Causal Condition

Before discussing the problems with NESS- and HP-causation, I present CNESS-causation [2]. As a first step, we define NESS-causation as the transitive closure of direct NESS-causation, which is how it was conceived by Wright [33]. In addition, we pay explicit attention to the path along which the causal influence is transmitted.

Definition 8 (**NESS).** $X = x$ NESS-causes $Y = y$ along a path p *w.r.t.* $(M, \vec{u})$ *if the values of the variables in* p *form a path of direct-NESS causes from* $X = x$ *to* $Y = y$.

The *Counterfactual* NESS definition (CNESS) takes the NESS definition and adds a subtle counterfactual difference-making condition: there should be a counterfactual value so that it would not NESS-cause the outcome along the same path as the actual value, nor along any subpath.

Definition 9 (**CNESS).** $X = x$ CNESS-causes $Y = y$ w.r.t. $(M, \vec{u})$ *if* $X = x$ *NESS-causes* $Y = y$ *along some path* p *w.r.t.* $(M, \vec{u})$ *and there exists a* $x' \in \mathcal{R}(X)$ *such that* $X = x'$ *does not NESS-cause* $Y = y$ *along any subpath* $p' \subseteq p$ *w.r.t.* $(M_{X \leftarrow x'}, \vec{u})$.

With all the definitions of causation at hand, I now motivate my choice for the CNESS definition by going over some well-chosen examples. We start with a case of Late Preemption.

Example 2 (**Late Preemption).** *We return to our two assassins, but this time* $Assassin_1$ *is slightly faster, so that their bullet kills Victim, who collapses and thereby dodges* $Assassin_2$*'s bullet.*

In this case $Assassin_2$ obviously did not cause Victim's death, and is thus not responsible for the outcome (despite the fact that their act itself is of course still blameworthy). BvH only allow variables for strategies and are thus unable to capture this result, since the asymmetry between both assassins is not a matter of strategy. As illustrated at length by Halpern [16], using causal models this poses no problem. The equation $V = BH_1 \vee BH_2$ expresses the fact that either bullet hitting Victim would be fatal; $BH_1 = A_1$ and $BH_2 = A_2 \wedge \neg BH_1$ captures the asymmetry between both assassins: $Assassin_2$'s bullet only hits Victim if $Assassin_1$'s bullet does not. In the context at hand, we have that $A_1 = A_2 = BH_1 = V = 1$, and $BH_2 = 0$. We now go through the various definitions to verify that $A_1 = 1$ NESS-causes, CNESS-causes, and HP-causes $V = 1$, whereas $A_1 = 1$ does not directly NESS-cause $V = 1$, thereby showing that the direct NESS definition is too simplistic.

We start by verifying that $A_1 = 1$ does not directly NESS-cause $V = 1$. By itself $A_1 = 1$ does not form a sufficient set for $V = 1$, for setting both of the BH variables to 0 guarantees that the Victim survives: $(M, \vec{u}) \models [A_1 \leftarrow 1, BH_1 \leftarrow 0, BH_2 \leftarrow 0]V = 0$. In fact, in this context, any sufficient set for $V = 1$ has to contain $BH_1 = 1$, yet BH_1 is sufficient for $V = 1$ all by itself. Thus $A_1 = 1$ is not a necessary member of any sufficient set for $V = 1$. Still, $A_1 = 1$ NESS-causes $V = 1$ along $p = \{A_1, BH_1, V\}$, because $A_1 = 1$ directly NESS-causes $BH_1 = 1$ and $BH_1 = 1$ directly NESS-causes $V = 1$.