eAppendix for "On the distinction between interaction and effect modification" by T.J. VanderWeele. Formal Identification Arguments for Joint and Conditional Causal Effects in Figure 4.

In this eAppendix, we show that for the causal DAG given in Figure 4, it is possible to identify the joint effects, $\mathbb{E}[D_{eq}]$, of E and Q on D and thus to assess interaction but that it is not possible to identify conditional causal effects of the form $\mathbb{E}[D_e|Q=q]$ and thus not in general possible to assess effect modification in Figure 4.

We first use Result 2 in Appendix 2 to show that the joint effects, $\mathbb{E}[D_{eq}]$, of E and Q on D are identified in the example represented by Figure 4. If, in Result 2, we choose $A_1 = E$, $A_2 = Q$ and $W = \varnothing$, we can see that all backdoor paths from E to D are blocked in the graph with the arrows into Q removed. Furthermore, if we select $V = X$ then we can easily verify that all backdoor paths from Q to D on the original graph are blocked by (E, X); the backdoor paths $Q - U_2 - E - D$ and $Q - U_2 - E - X - D$ are both blocked by E; the backdoor path $Q - U_1 - X - D$ is blocked by X; and the backdoor path $Q - U_1 - X - E - D$ is not blocked by X (since X is a collider on this path) but it is nevertheless blocked by E. Thus we can apply Result 2 to Figure 4 to identify the joint effects, $\mathbb{E}[D_{eq}]$, of E and Q on D and thus to assess interaction between the effects of E and Q on D.

We now show that quantities of the form $\mathbb{E}[D_e|Q=q]$ are not identified in causal DAG given in Figure 4. The argument is subtle and uses a number of technical results concerning causal DAGs.[28,31] First we note that there is a backdoor path from Q to D in the graph corresponding to Figure 4 with the node E removed, namely $Q - U_1 - X - D$; from Theorem 6 of Shpitser and Pearl[28] we have that $P(D_e|Q=q)$ is identified if and only if $P(D_e, Q_e)$ is identified. Since E has no effect on Q in Figure 4, it follows that $P(D_e|Q=q)$ is identified if and only if $P(D_e)$ is identified. Now, in Figure 4, there is path from E to X which consists entirely of consecutive confounding arcs, namely $E - U_2 - Q$ and $Q - U_1 - X$; X is a child of E and from Theorem 3 of Tian and Pearl[31] it follows that $P(D_e)$ and thus that $P(D_e|Q = q)$ is not identified.

Intuitively, one might reason that if we are interested in estimating the effect of E on D conditional on Q we must use data on E, Q and D. However, if one controls for X, then control is being made for an effect of E and this will bias the estimate. If control is not made for X then there is an unblocked backdoor path from E to D, namely, $E- U_2-Q-U_1-X-D$ (note that this path is unblocked because Q is a collider on this path and one is conditioning on Q). The situation may seem analogous to the classical time-dependent confounding issue (that marginal structural models handle) in which a confounder of a subsequent exposure is on the causal pathway between prior exposure and the outcome. However, in Figure 4, unlike in the time-dependent confounding case, marginal structural models cannot help in the identification of the effect of interest, $\mathbb{E}[D_e|Q = q]$. The argument given above using the results of Shpitser and Pearl[28] and Tian and Pearl[31] demonstrates that the $P(D_e|Q = q)$ is not identified; no method of adjustment can be used to identify $\mathbb{E}[D_e|Q = q]$. The distinction is that in the time-dependent confounding case, data is available to identify the causal effects of interest but simple adjustment approaches like regression and stratification do not suffice; inverse-probability-of-treatment weighting techniques are needed. In Figure 4 the issue does not concern the method of adjustment but rather the fact that the data available are insufficient to identify the conditional causal effect of interest.