Why do add the interaction and the main effect

For x_1 denoting political ideology and x_2 representing the topic either relevant ($x_2 = 1$) or irrelevant to the ($x_2 = 0$) to the comment.

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon$$

$$\begin{aligned} \frac{\partial}{\partial x_1} y &= \beta_1 + \beta_3 x_2 \\ &= \begin{cases} \beta_1 + \beta_3 & \text{if } x_2 = 1 \\ \beta_1 & \text{if } x_2 = 0 \end{cases} \end{aligned}$$

In the present case $\beta_1 < 0$, meaning that more conservative individuals perceive content as more toxic on average (across topics). We are interesting in when the interaction effect is "undoing" the main effect – i.e. $\beta_1 + \beta_3 > 0$