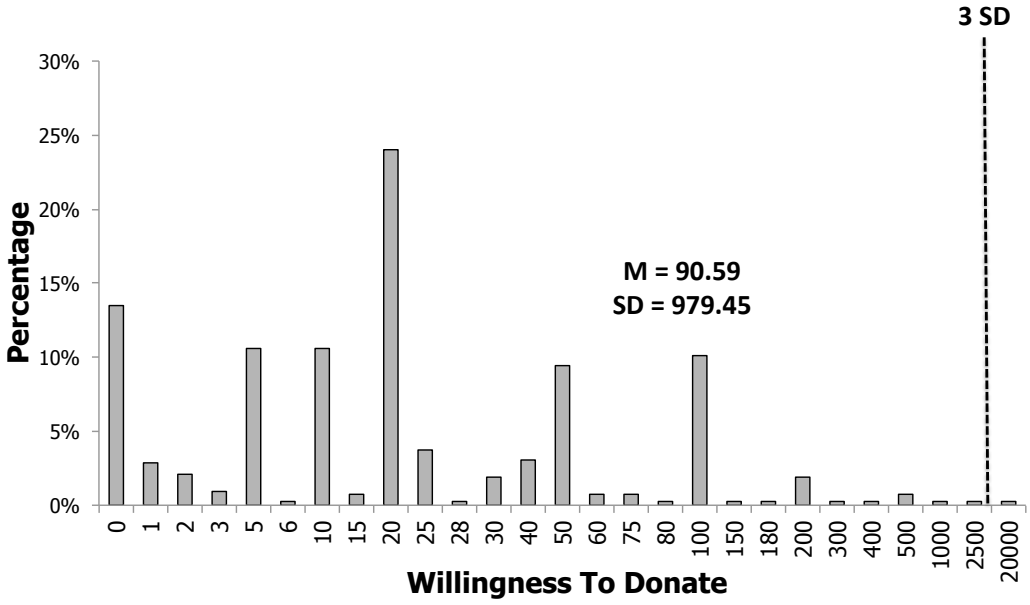
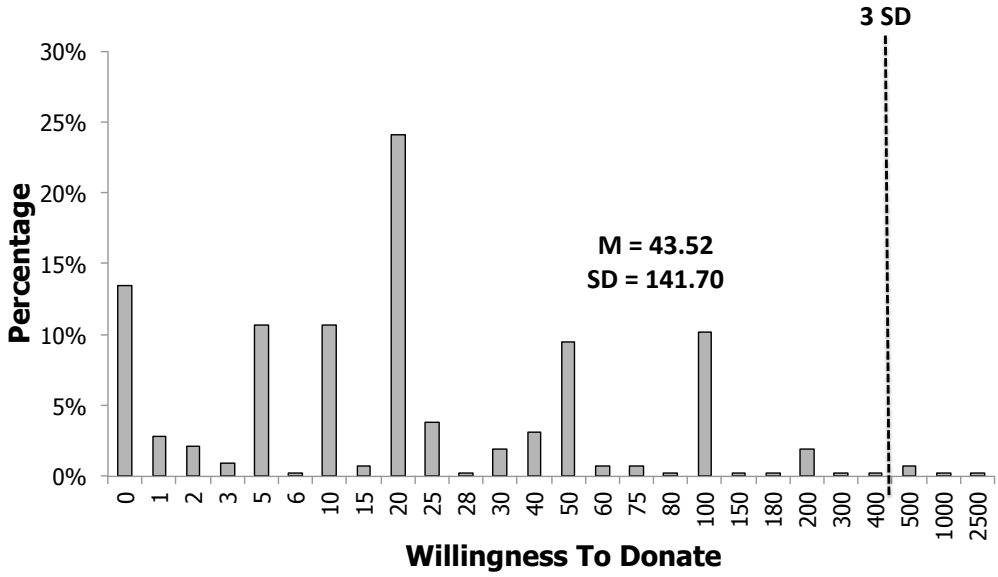
Asking people how much they'd donate to 20 needy schoolchildren can produce outlier responses. For example, this is the distribution of donations in my exact replication:



Hsee et al. removed outliers greater than 3 SD from the overall mean. If I did that here, then I'd remove all donations greater than $2,917.43 and retain all others. Thus, I'd remove only the $20,000 observation.

But what if that obviously outlandish outlier had not been in our dataset? In that case, I'd remove the five donations greater than $454.61.



Because one very large outlier can dramatically affect the mean and SD, it can dramatically affect which responses get excluded when the exclusion criterion is SDs from the mean. In my exact replication, the results are *not* robust to using the $20,000 observation to draw the exclusion line (opposite direction, p=.623); they emerged only when I first removed that item, and then removed responses greater than 3 SD from the rest of the sample's mean (donations>$454.61).

Given that the $20,000 response is simultaneously completely absurd and extremely influential, I think it makes sense to remove it prior to setting the 3 SD cutoff. Still, it might be better to identify outliers in a way that is not so susceptible to a single extreme observation. To their credit, Hsee et al. also report that their results show the "same pattern" when Winsorizing the data so that all values greater than the 95th percentile were assigned a value equal to the 95th percentile. When I did that in my replications, Hsee et al.'s results replicate (p=.066 in the near replication and p=.001 in the exact replication).