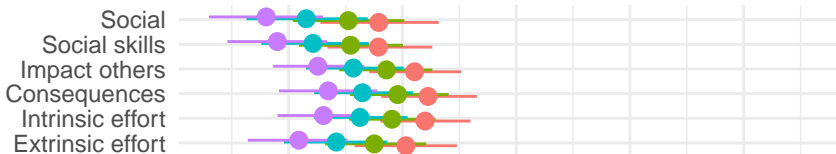


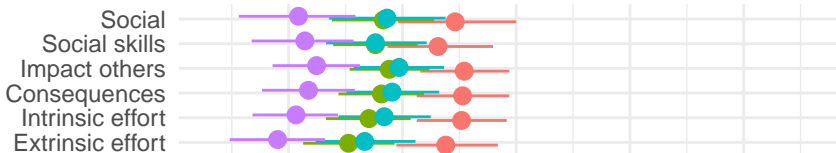
## Competent



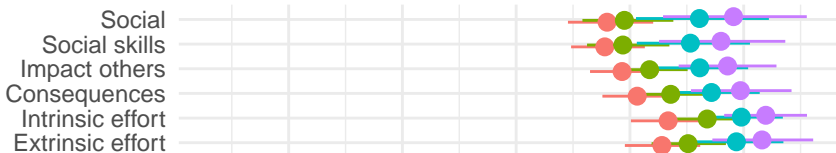
## Warm



## Moral



## Lazy



## Trustworthy



- Tool-Honest
- Tool-Deception
- Full-Honest
- Full-Deception

-1.0 -0.5 0.0 0.5 1.0

Estimate