Simon Zhuang and Dylan Hadfield-Menell. 2021. Consequences of Misaligned AI. http://arxiv.org/abs/2102.03896.