Appendix I
Charness-Rabin model

We use definitions that stem from the model of Charness and Rabin (2002) who consider the following simple formulation of the preferences of $self$:

$$u_s(\pi_s, \pi_o) \equiv (1 - \rho r - \sigma s)\pi_s + (\rho r + \sigma s)\pi_o,$$

where $r = 1$ ($s = 1$) if $\pi_s > \pi_o$ ($\pi_s < \pi_o$) and zero otherwise. Notice that proportionally increasing ρ and σ indicates a decrease in self-interestedness whereas increasing the ratio ρ/σ indicates an increase in concerns for increasing aggregate payoffs rather than reducing differences in payoffs. Thus, the parameters ρ and σ allow for a range of different distributional preferences:

(i) *competitive* preferences ($\sigma \leq \rho < 0$), where utility increases in the difference $\pi_s - \pi_o$, are consistent only with the competitive allocation $\pi^c = (\pi_s^s, 0)$;

(ii) *narrow self-interest* or *selfish* preferences ($\sigma = \rho = 0$), where utility depends only on π_s, are consistent with any allocation π where $\pi_s = \pi_s^s$;

(iii) *difference aversion* preferences ($\sigma < 0 < \rho < 1$), where utility is increasing in π_s and decreasing in the difference $\pi_s - \pi_o$, are generally consistent with the allocations π^s and π^e if $\pi_s^e = \pi_s^o$;

(iv) *social welfare* preferences ($0 < \sigma \leq \rho \leq 1$), where utility is increasing in both π_s and π_o, are only consistent with π^s and π^o.

To provide a clearer intuition, Figure AI1 illustrates difference aversion and social welfare preferences and depicts the range of solutions when $\pi^e \in \Pi^3$. A typical indifference curve for difference averse preferences is represented in the left panel ($MRS_{os} > 0$ for $\pi_s < \pi_o$) and for social welfare preferences in the right panel ($MRS_{os} < 0$ for $\pi_s < \pi_o$). In these cases, the difference aversion optimum is π^s or π^e whereas the social-welfare optimum is π^s or π^o. Also notice that many allocations are not consistent with any of the above prototypical preferences. For example, any allocation $\pi \in \Pi^3$ is not consistent with any of these preferences unless $\pi = \pi^e$.

[Figure AI1 here]

Figure AI1: An example of the preferences of Charness and Rabin (2002)
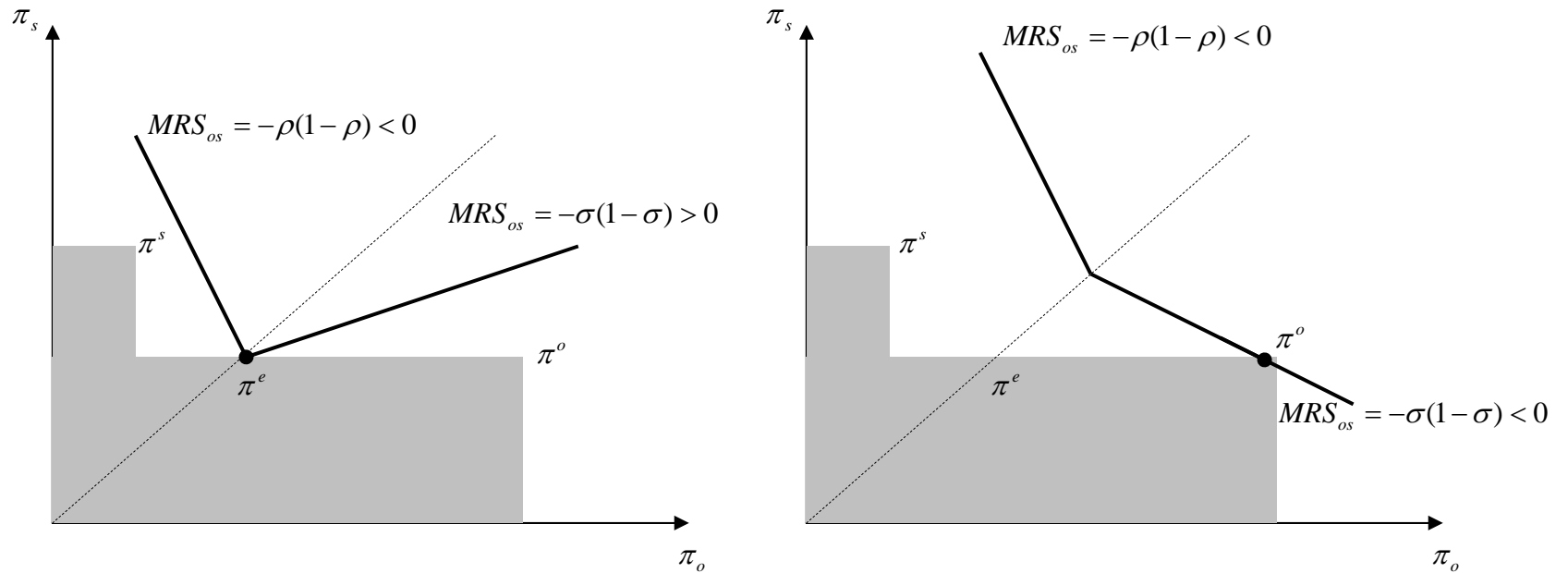
Instances of social preferences and the range of solutions when $\pi^e \in \prod^3$. A typical indifference curve of a difference aversion function is represented in the left panel and of a social-welfare function in the right panel. The difference aversion optimum is π^e whereas the social-welfare optimum is π^o.