R2 & R5: "Threshold policies are optimal". We thank **R2**, **R5** for the reference (Corbett-Davies et al., 2017). Our Thm 1 is indeed similar. We will be sure to cite this work and explain the relationship with ours.

R2 & R5: "COMPAS experiments". We share the same reservation in using this dataset and lending validity to Condition 1A, even though our purpose was only to show the flexibility of our framework. As an alternative, we've run experiments on COMPAS under all conditions (1A-D). Table 1 shows Prop 1 holds under 1A-B, no oscillation under 1B-C, and more uncertainty under 1C-D, which is discussed in Appendix F. We will replace the original version with these results, or, if this still does not address reviewers' concerns, we are fine with removing this experiment altogether.

Table 1: osi*/osi$_H$/osi$_L$ is the percentage that oscillation occurs among 125 set of different transitions under policy $\texttt{UN}^*/\texttt{UN}_{\theta_H}/\texttt{UN}_{\theta_L}$. Among transitions that lead to stable equilibrium, Col 2/Col 3 shows the percentage that $\texttt{UN}_{\theta_H}$/ $\texttt{UN}_{\theta_L}$ results in lower recidivism compared with $\texttt{UN}^*$.

	$\widehat{\alpha}_{\theta_H} < \widehat{\alpha}^*$	$\widehat{\alpha}_{\theta_L} < \widehat{\alpha}^*$	osi*	osi$_H$	osi$_L$
A	0	1	0.29	0.12	0.36
B	0.99	0.01	0	0	0
C	0.37	0.28	0	0	0
D	0.79	0.63	0.06	0	0.13

R2 & R3: "1D feature space". We thank **R2** and **R3** for suggestions on notations. To clarify, our work is not limited to 1D feature space (line 115-116); the generalization to $\mathbb{R}^d$ is in Appendix E.

R2, R3 & R5: "Interventions". **@R2** sensitivity: This is one of our main findings (lines 7-9) and reflects broader challenges in designing a fair policy: the effect of intervention highly depends on problem parameters, and the same intervention may lead to contrarian results as parameters change. **@R3** Trade-off: Prop. 1 improves $\widehat{\alpha}^s$ by sacrificing instant utility; Prop. 2 achieves equality but violates static fairness and sacrifices instant utility. Note that the sacrifice in instant utility in both cases may actually result in improved long-term total utility (line 283). While T^s_{yd} is a group property, it is controllable through community-level interventions (Prop. 3) such as social support (subsidy, training, etc.) to sub-populations (line 294). **@R5** usage: Intervention remains feasible even without knowing the true qualification (of the rejected population); e.g., supporting all those rejected (or accepted) would increase both T_{1d} and T_{0d}.

R2: "Related work in economics". The model studied in (Coate and Loury, 1993) is more relevant to [33], where people manipulate their qualifications and groups have identical feature distribution and response to the policy. We will add this comparison and more related works in economics. "Limitations on group-specific policies". In some cases the use of sensitive attribute is allowed (e.g., per ECOA Regulation B, *age* can be used in lending in the US). Nonetheless, we will clarify that group-specific policies may not be generally applicable.

R3:"Stability of unique equilibrium". It is true there can be oscillation under conditions in Thm.3, as they only guarantee uniqueness but not stability. We discuss stability in Appendix F (line 662-677), and will clarify this in the main body. "Harm/benefit of fairness if natural equality is not broken". This is examined in Thm.4 (line 218-219): equality is violated/maintained if distributions are different/same."Scenarios under natural inequality".Under our model two groups can be different in transition or/and feature distribution. Natural inequality arises when either one or both differ across two groups. We thus regard these as two sources of inequality, consider them separately by fixing one and varying the other (two scenarios we studied), and examine whether fairness constraints can address inequality caused by each (line 228-230). "When DP flips the advantaged group". In this case the gap $|\widehat{\alpha}^a_{\texttt{DP}} - \widehat{\alpha}^b_{\texttt{DP}}|$ highly depends on feature distribution. Empirical results (Table 3, Appendix A) show DP can reduce this gap (mitigate inequality): $\widehat{\alpha}^b_{\texttt{DP}} - \widehat{\alpha}^a_{\texttt{DP}} < \widehat{\alpha}^a_{\texttt{UN}} - \widehat{\alpha}^b_{\texttt{UN}}$.

R3 & R5: "Markov dynamics". As long as an appropriate "state" (sufficient statistics) can be identified, the Markov assumption holds; this in the worst case would be the entire history which would indeed be undesirable. In practice historical information is often summarized into a (pseudo) sufficient statistics to enable tractable decision making; e.g., lending decisions rely on the entire history only through summaries such as the credit score or a set of scores, which can be regarded as the state in a Markov process. **@R3** linearity: Although individual action doesn't depend on α^s_t, we note Eqn. (4) is *not* linear in qualification α^s_t, as g^{1s}_t and g^{0s}_t are functions of policy π^s_t, which is nonlinear in α^s_t.

R3 & R5: "Justification on modeling choices". **@R3**: In addition to current explanations and examples (line 162-165 on transitions; line 234-236, 254-255 on two scenarios under natural inequality), we will further strengthen the motivation for these modeling choices. **@R5** POMDP: We think POMDP is a reasonable framework to capture the sequential nature of the problem and the fact the true qualifications Y are unobservable to the decision maker, and the decision is based on X and a belief state on Y. A (PO)MDP approach has been motivated and used in similar studies [9,23]. In particular, [9] shows that although many works on fairness didn't explicitly use (PO)MDP to model dynamics, they can all be cast into the standard framework of (PO)MDP, such as works on lending [32,36], college admission [22,28], attention allocation [11], etc. The studies on the scenario mentioned by **R5** (outcomes are observed only under positive decisions) are orthogonal to our work, which we introduce and discuss in Appendix C (line 596-598).

R5: "Conditions 1A-D". Under our POMDP framework (Fig. 1), dynamics of α_t follow Eqn. (4) and 1A-1D actually capture all possibilities including the case mentioned by the reviewer. Specifically, if $\text{avg}(T_{01} + T_{11}) \leq$ (resp. $\geq$) $\text{avg}(T_{00} + T_{10})$ holds, then either 1A (resp. 1B) or 1C or 1D must hold. "Explain conditions in Thm. 3". This can be considered as a weaker version of the Lipschitz condition. More discussion is in Appendix F (line 662-677).