We thank the reviewers sincerely for their valuable feedback, which has provided us with an achievable plan to improve the clarity and impact of the paper with a revision as outlined below. We were particularly encouraged by the predominantly positive and supportive remarks (e.g., R1: "*The result of fair resource allocation in continuous time is novel and interesting.*" and R4: "*This is a solid technical work which provides (1) characterization for optimal policy as well as (2) an online algorithm with a guaranteed regret bound. The use of renewal theory is interesting.*") In the following, we present our responses to the specific questions of the reviewers.

Q1. Clarification of the system model and objective (R2) We would like to thank Reviewer 2 for pointing this out. We will enhance the presentation of the system model to avoid confusions and improve clarity. We consider a controller (or decision maker) that allocates tasks sequentially to one of K groups. A group is analogous to an *arm* in the stochastic bandit setting, where arm pulls from a given arm yield i.i.d. rewards. Analogously, each task completion (by an individual) from a given group yields a random reward. We assume these completion times and rewards are i.i.d. within each group. The decision-making process continues until the total time spent exceeds a given time budget B, and the objective of the controller is to maximize the utility subject to this budget constraint. Using different utility functions, we are able to learn a fair allocation of the budget B across groups with different notions of fairness, e.g., proportional fairness. Since the controller does not have any prior statistical knowledge on the completion times and rewards for any groups, learning and utility maximization need to be performed simultaneously. Our i.i.d. [reward, completion time] assumption implies a statistical symmetry between individuals (*arm pulls*) within a group (*arm*). Therefore, allocation of individuals within groups is *not* a part of the decision-making in our setting.

Q2. Why do we assume group-dependent and i.i.d [reward, completion time] pairs? (R1, R2) We would like to thank the reviewers for this important question. We will further justify the statistical assumptions in the final version of the paper. In many cases, the individuals within a group exhibit random but *statistically similar* (i.e., independent and identically distributed) performance as a consequence of their common demographic (e.g., resources for training) background. We agree that individuals within a given group may have very different skill sets, and we reflect this as random and potentially *heavy-tailed* completion times for each individual, which implies that there can be frequent statistical outliers within a group that deviate from the mean tremendously depending on the tail. We also note that our approach in this paper can be utilized to develop algorithms for different and more complicated statistical models, such as time-varying (e.g., Markov-modulated) completion time and reward statistics.

Q3. Why do we assume delayed full information feedback? (R3, R4) As noted by Reviewer 1, a key contribution of this work is to introduce the powerful Lyapunov drift methods into the fair online learning framework. This is a challenging task because of the interaction between the state dynamics and the empirical estimates. Thus, as Reviewer 3 points out, the delayed full feedback assumption facilitates the analysis, and provides a methodological basis for other more complicated feedback models (analogous to the *expert advice* in bandits). We will further justify in the paper that there are applications for which the delayed full-feedback model is accurate. For example, in server allocation and two-sided matching markets with online reviews, the completion time and reward realizations (or quality estimates through interpolation) are shared among decision-makers after a delay. Thus, the feedback of an unchosen group for the n-th task is available only after τ tasks. Nevertheless, we agree that there are also many applications for which full-information feedback is unavailable, which we leave as an open question. We believe this question is highly non-trivial since state-dependent online exploration will be crucial in partial feedback models, which makes the analysis of the Lyapunov drift methodology considerably harder because of the complex interactions between the state and statistical estimate for each decision.

Q4. Extended Performance Evaluations of the `OLUM` Algorithm (R1, R4) In the final version of the paper, we will include extended numerical investigations. Our updated set of experiments will address the motivating examples of server allocation in cloud computing systems and contractual hiring in online freelancing platforms, for which accurate power-law statistical models are available in the literature (e.g., [36], [37]) , and include: (i) the impact of the alpha fairness parameter on the performance of the `OLUM` Algorithm, (ii) the case of multiple groups (as suggested by Reviewer 4), and (iii) the impact of the tail distributions on the resource allocation. By these investigations, we will observe how regret scales with the number of groups and statistical heterogeneity in practice.

Q5. Writing and Clarifications (R1, R3) We will expand the technical sections in the paper to provide more insight on the theoretical novelties that we introduce. In particular, we will emphasize the novel challenges and our solutions in developing a learning algorithm based on empirical Lyapunov-drift estimates as suggested by Reviewer 1, this will also highlight the difference with [19], which uses an exact Lyapunov drift that requires complete statistical knowledge. Also, we will add more insights on our renewal theory-based approach to utility maximization in continuous time (as suggested by R1), which is a less-studied problem in the literature. We will also clarify the impact of R_{max} in the theorem statement, which was previously discussed in the proof (as mentioned by R3).