We are glad to see that the reviewers found our work interesting, relevant, and sound. We also appreciate the feedback provided, which will be incorporated to improve the manuscript. Below, we address each reviewer's comments.

Reviewer #1 [A. *Term 'Simplicity'*]: The term simplicity in this context is about having fewer actions and contexts while guaranteeing the same optimal performance. Without simplicity, additional parameters that are irrelevant will need to be estimated during the exploration stage, and will possibly slow down the whole learning process. To ground this discussion, we contrast the MPSes of Fig. 3c and Fig. 3d. In particular, the MPS in Fig. 3d is $\pi(X_1|C)\pi(X_2|C)$ while the one in Fig. 3c is $\pi(X_1|C)\pi(X_2|C, X_1)$. While both MPSes coincide in terms of $\pi(X_1|C)$, the second component of MPS 3d — $\pi(X_2|C)$ — is "simpler" (i.e., lower dimensional) than that of MPS 3c, $\pi(X_2|C, X_1)$. We'll try to improve the discussion in the paper, thanks. [B. *Put in the context of RL*]: This work focuses on policies over a general causal graph, which traditional RL framework lacks. Actions and contexts are typically fixed a priori in RL. We tried to emphasize such differences through the introductory example. For instance, given a traditional MDP structure, $\{\pi(x_t|s_t)\}_t$ is considered the only POMPS. [C. *References*]: Thanks for the suggestions, they appear to be interesting and relevant, will check. [D. *Identifiaiblity issues*]: The focus of the paper was on characterizing mixed policies given that the agent can act (i.e., online learning). In other words, the focus is not in doing off-policy evaluation and using the data collected by another agent, which is when identifiability comes into play. [E. *Term 'DAG' and 'Mixed Graphs'*]: We meant by a 'DAG with latent variables' (also called a causal diagram), which is represented as an ADMG under latent projection. We follow Pearl's notation, but will add a clarification note about it, thanks. [F. *Explicitizing bidirected edges to UCs in Thm. 3*]: Thm. 3 makes use of a specific instantiation of the unobserved confounders for the sake of reduction. However, it holds true irrespective of such instantiation. In fact, one can employ $\mathbf{u}_X$ (which also includes variable-specific exogenous variables that can be effectively ignorable, see line 94) for $X \in \mathbf{X}'$. For concreteness, see Fig. 7b, where the two bidirected edges are mapped to $\{U_1, U_2\}$. Whenever they appear in the derivation, $\mathbf{U}_{X_1} = \{U_1, U_2\}$ can be used, $\mathbf{U}_{X_2} = \{U_1\}$, or their union, without relying on a specific instantiation. Having said that, that's a good point, we will try to make it more direct and explicit in the manuscript. [G. *Term 'Observe' and 'Listens to'*]: The terms are used to describe 'being used as a context' (i.e., contextualized). There are multiple similar terms (observe, listen, see, and contextualize). We borrow the terminology from Pearl but will make their meanings clearer to avoid any confusion. [H. *Acyclicity in the definition of MPS*]: In the definition of MPS, $\mathcal{G}_{\mathcal{S}}$ being acyclic is the condition we desire (i.e., a desideratum). In other words, we did not consider a set of actions and their contexts that would create a cycle in its induced graph $\mathcal{G}_{\mathcal{S}}$. If you can further specify why the definition is unclear, we will be able to address the issue better. Thanks for letting us know recent results about acyclicity-inducing policies on a more general class of graphs. We will look into the connection.

Reviewer #2 [I. *Practicality of the results*]: As you have mentioned, we made an interesting connection between causality and RL by providing theoretical insights. We expect that, as we better understand such connections, our theoretical results will guide the design of practical algorithms or applications. [J. *Complexity Analysis*]: We consider studying efficient algorithmic characterizations of Thm. 2 and 3 as important research directions. At the moment, these theorems provide graphical understandings of redundancy and optimality. Hence, we do not have algorithmic characterizations for Thm. 2 and 3. [K. *Cost of calculation*]: Both the time steps till the convergence and time for calculation seem two important aspects when considering time/computational cost. We will elaborate in the paper the implications of the cost of calculation in evaluating the performance of agents. [L. *D-Separation*]: Thm. 2 primarily focuses on whether the variations of some of contexts are ignorable by examining their fixability. In doing so, partly, Thm. 2 utilizes the property of d-separation with deterministic relationships. Thm. 2 being a sufficient condition for non-NRO is orthogonal to d-separation being complete for conditional independence.

Reviewer #3 [M. *Accessibility to ML audience*]: We tried to utilize figures and examples in most cases to be accessible. We will be able to make the paper more accessible given one additional page for the accepted paper. [N. *Causal graph assumption*]: It is assumed but also can be (partially) inferred from existing data or through interactions. Studying under the availability of a causal graph is a necessary step towards better understanding what agents can/should do with policies when the agent can only access to partial information about the underlying environment. [O. *Reorganizing preliminaries/Appendix A*]: We are planning to incorporate the essence of Appendix A into introduction so as to better motivate the paper. [P. *Term 'listening'*]: The term 'listening' is used to carry the meaning of the agent being 'actively engaging in observing' variables to determine actions (i.e., to use as a context). Please read also [G] in Reviewer 1.

Reviewer #4 [Q. *Practicality of the results*]: Please check out [I] in Reviewer 2. [R. *'Surprising results' in the abstract*]: We intended to emphasize that a policy trying to intervene more (or all) variables blindly, even with utilizing contexts, can be failed to converge incurring regret all the times. People often believe that intervening more variables always leads to a better outcome, which is not always the case. (We also agree that 7e is another surprising result since non-ancestors of Y in $\mathcal{G}$ are considered irrelevant.) [S. *Term 'optimality' in the introduction*]: Thanks for catching our mistranslation of an expression $\neg\forall_{\mathcal{M}\sim\mathcal{G}}\exists_{\mathcal{S}'\neq\mathcal{S}}\mu^*_{\mathcal{S}} \leq \mu^*_{\mathcal{S}'}$, which is the double-negation of Def. 5 (roughly) $\exists_{\mathcal{M}\sim\mathcal{G}}\forall_{\mathcal{S}'\neq\mathcal{S}}\mu^*_{\mathcal{S}} > \mu^*_{\mathcal{S}'}$. We are planning to revise the introduction example with more formal notation (redundancy and optimality) to avoid any confusion. A relevant response is also in [O]. [T. *Thm. 2 is involved*]: We will improve the presentation of Thm. 2 by incorporating additional figures such as Fig. 15 in the Appendix. Thank you for the suggestion.